SOME ASPECTS OF STATISTICAL INFERENCE

IN THE LINEAR REGRESSION MODEL

A thesis presented for the degree

of Doctor of Philosophy

in Economics

at

the University of Canterbury

Christchurch, New Zealand

by

M.L. King

November 1979

CONTENTS

CHAPTER PAGE

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This thesis considers two aspects of statistical inference associated with the linear regression model set in an economic context. The implications of replacing the conventional normality assumption with the broader assumption that the disturbances follow an elliptically symmetric distribution, are investigated, and three features of the problem of detecting serial correlation in elliptically symmetric disturbances, are studied.

An examination of the conventional justification of the normality assumption in econometrics, conducted in Chapter 2, provides motivation for the study of regression analysis under the elliptical symmetry assumption.

In Chapters 4 and 5, properties of estimators and tests associated with the linear regression model are investigated, assuming elliptically symmetric disturbances. This broadening of the normality assumption is found to have few practical consequences for classical regression analysis. The usual least squares estimators are shown to satisfy a stringent optimality property. Conditions are determined for weak consistency and for strong consistency of these estimators. Distributions of statistics invariant to the disturbances' scale are found to be unaffected by the broadening of the normality assumption, while the distributions of arbitrary statistics can be viewed as mixtures of their distributions for different scales of the disturbances under normality. The implications of these results for hypothesis testing, are explored.

Chapter 6 attempts to find an "optimal" test for first-order autoregressive disturbances based on LUS residuals. A test, which is optimal for certain design matrices, is constructed and shown to be the Abrahamse-Koerts test.

Durbin and Watson's and Kramer's procedures for applying the Durbin-Watson test to a regression equation without an intercept, are compared in Chapter 7. While being susceptible to pre-test bias, Kramer's procedure is found to have superior power.

Chapter 8 considers the problem of detecting first-order moving average disturbances. The Durbin-Watson test is shown to be approximately locally best invariant. A new test is proposed and found to have useful power properties.

CHAPTER 1

INTRODUCTION

1. INTRODUCTORY COMMENTS

The underlying philosophy of this thesis is provided by the following two fundamental propositions:

(i) The usefulness of any theory of statistical inference depends on, amongst other things, the generality of the assumptions upon which the theory is based.

(ii) With respect to a problem of hypothesis testing, it is desirable that full regard be given to the power properties of the resultant test when the choice of test procedure is made.

This thesis makes a number of contributions, in the spirit of (i) and (ii), to the classical theory of statistical inference associated with the linear regression model set in an economic context.

There is a growing body of evidence which suggests that the assumption of normally distributed disturbances, that underlies much of the classical theory, may often be an unrealistic one. A review of this evidence, together with a critical examination of the conventional justification of normality, leads one to the conclusion that this traditional assumption will frequently be violated in practice. This provides motivation for wanting to consider a theory of inference based on a wider class of disturbance distributions than that of multivariate normal.

The major portion of the thesis is therefore devoted to determining the implications, for the classical theory of regression analysis, of replacing the normality assumption with the wider, and perhaps more acceptable assumption that the regression disturbances follow an elliptically symmetric distribution.

Some of the results of this inquiry are used in the second part of the thesis where three aspects of the problem of testing for serial correlation in elliptically symmetric regression disturbances are investigated in the spirit of proposition (ii). In this part of the thesis, an attempt is made to find an "optimal" exact test for first-order autoregressive disturbances, the question of applying the Durbin-Watson bounds test to a regression fitted throught the origin is discussed, and the potential of the Durbin-Watson test as a test for first-order moving average disturbances is explored.

2. AN OUTLINE OF THE THESIS

Chapter 2 briefly reviews the controversy surrounding the "Gaussian law of error" as a theory of measurement error, and then critically examines the validity of the normality assumption in the linear regression model, set in an economic context, from both a theoretical and an empirical point of view.

Definitions of spherically symmetric and elliptically symmetric distributions are given in Chapter 3. This is followed by a survey of the literature concerned with the properties of these distributions. The chapter closes with a summary of those properties which are used elsewhere in the thesis.

The subsequent two chapters investigate various aspects of statistical inference in the linear regression model with elliptically symmetric disturbances. In Chapter 4, a number of properties of the ordinary least squares and the generalized least squares estimators of regression coefficients are established. Conditions are given for weak consistency and strong consistency of both estimators, and their asymptotic distributions are discussed. The generalized least squares estimator is shown to satisfy a stringent, and intuitively desirable, optimality property, as well as being the maximum likelihood estimator.[1]

The size and small sample power properties of the numerous statistical tests associated with the classical linear regression model are investigated in Chapter 5.[2] The results established in this chapter also enable the distributions of regression parameter estimators to be related to their distributions for normally distributed disturbances. The first half of the chapter considers the small sample distributions of regression statistics which are invariant to the scale of the disturbances, while the latter half deals with the distributions of arbitrary statistics for a restricted class of elliptically symmetric disturbances.

Chapter 6, 7 and 8 each consider different aspects of the problem of testing for serial correlation in elliptically symmetric regression disturbances.

---

1. For a slightly restricted class of elliptically symmetric disturbance distributions.

2. A paper presenting some of the author's preliminary results concerning the optimality of various tests for serial correlation in elliptically symmetric regression disturbances, has been accepted for publication in the *Annals of Statistics*. The results reported in Chapter 5 are more general in that they apply to a broader class of elliptically symmetric distributions than that considered in this earlier study.

Chapter 6 begins by briefly reviewing that part of the large body of econometric literature on the subject of testing for serial correlation which is concerned with the construction of exact tests as alternatives to the Durbin-Watson bounds test. A number of empirical investigations suggest that the small sample properties of most of these tests are distinctly inferior to those of the Durbin-Watson test. A possible explanation is that proper regard was not given to the potential power of these exact tests in their construction. Chapter 6 reports an attempt to find the "optimal" exact test based on linear unbiased residuals with a scalar covariance matrix.

Chapter 7 considers the problem of applying the Durbin-Watson bounds test to a regression equation fitted through the origin. The small sample power properties of the procedure suggested by Durbin and Watson (1951) and an alternative procedure proposed by Kramer (1971) are compared. The possibility of pre-test bias in the latter is discussed and Kramer's procedure is extended in order to test for negative autocorrelation.

The potential of the Durbin-Watson test, as a test for first-order moving average disturbances, is investigated in Chapter 8. It is found to be an approximately locally best invariant test. A new exact test for first-order moving average disturbances is proposed and the small sample power properties of the two tests are compared for selected design matrices.

Concluding remarks are made in Chapter 9.

3. CONVENTIONS USED IN THE THESIS

Throughout this thesis, upper case characters are used to denote matrices or sets, while lower case characters are reserved for scalars and vectors. In keeping with current practice in econometrics, no attempt is made, by the use of symbols, to distinguish between a random variable or vector and the value taken by that random variable or vector.

The term "orthogonal" has two meanings depending on its context. An *orthogonal* matrix is defined as a non-singular matrix whose transpose is its inverse. On the other hand, a set of non-zero vectors, $\{x_1,\ldots,x_n\}$, is defined to be *orthogonal* if

$$x_i'x_j = 0, \qquad \text{for } i,j=1,\ldots,n, i\neq j.$$

An orthogonal set of vectors is orthonormal, if and only if all vectors have unit length; i.e., if and only if

$$x_i'x_i = 1, \qquad i=1,\ldots,n.$$

CHAPTER 2

A CRITICAL EXAMINATION OF THE NORMALITY ASSUMPTION IN THE LINEAR REGRESSION MODEL

1. INTRODUCTION

The assumption of normally distributed disturbances underlies much of the classical theory of statistical inference associated with the linear regression model. The aim of this chapter is to critically examine the relevance of this assumption for the linear regression model set in an economic context.

Over the past century, the role of the normal distribution in statistics has increasingly become a matter of controversy. This controversy is briefly reviewed in Section 2 and an attempt is made to clarify the issues involved. Historically, the emphasis has been on the role of the normality assumption in problems in which the stochastic variability is solely due to measurement errors; the supposition that such errors are normally distributed being known as the Gaussian law of error. Section 2 closes with a discussion of the theoretical and empirical objections to the Gaussian law of error.

These objections do not necessarily apply to econometric models because the stochastic variability in such models typically has other causes besides measurement error. Section 3 critically examines the conventional justification for the normality assumption in econometrics. A number of theoretical objections are discussed and the empirical evidence is reviewed. Some concluding remarks are made in Section 4.

2. THE GAUSSIAN LAW OF ERROR

Although the normal distribution is often named after Gauss, it was first identified by De Moivre in 1733 as the limiting form of the binomial distribution. Without reference to De Moivre's work, Gauss (1809) rediscovered the normal distribution while formulating his theory of measurement error. Gauss started from the premise that when a number of equally good observations of an unknown quantity, x, are given, the mean of the observations is generally accepted as the most accurate estimate of x. Assuming independent observations, he then found the only error distribution which satisfied his initial premise was the normal distribution. In other words, Gauss constructed the normal distribution to best suit the sample mean as an estimator of x.

Despite this tenuous line of reasoning,[1] the case for the Gaussian law of error was strengthened by Laplace (1812) who formulated the first (incomplete) statement of the Central Limit theorem - that under general conditions on the parent distribution, the distribution of the mean of a random sample tends to normality as the sample size increases. Unfortunately, Laplace and a number of later writers tended to overstate the generality of what had been proved. Statements were made in support of the Gaussian law of error implicitly assuming that the parent distribution of a random sample converges to a normal distribution as the sample size increases.[2] An example of such an implicit assumption may be found in

---

1. Evidence which demonstrates that the sample mean was not universally accepted in Gauss's time may be found in survey articles by Huber (1972) and Harter (1974a). Gauss, himself, admitted in 1823 that his earlier reasoning was not entirely satisfactory.

2. See Harter (1974a, pp. 155-6).

Laplace's (1812) claim[3] that for Gauss's problem, the best method of estimating x depends on the particular form of the law of error when the number of observations is small, but is the arithmetic mean when the sample size is large.

About this time, Poisson (1824) discovered that for the Cauchy distribution, the mean of a random sample follows the same distribution law as the parent distribution. Hagen (1837) and Bessel (1838) made a significant contribution to the understanding of the conditions required for Gaussian errors by developing their "Hypothesis of Elementary Errors". Their starting postulate was that the total error committed in any physical measurement is the sum of a large number of mutually independent elementary errors, each with the same absolute magnitude and equal probabilities of being positive or negative. The Central Limit theorem then implies that the total error is approximately normally distributed. Both Hagen and Bessel warned against the blind acceptance of the Gaussian law or error.

Despite such warnings and Poisson's discovery, for the greater part of the nineteenth century the Gaussian law of error enjoyed almost universal acceptance. The attitude of this era is perhaps best summed up in the following remark which Poincaré (1912) attributes to Lippman:

> "Everybody believes in the law of errors, the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is an experimental fact."

Cramér (1946, p.231) writes of the years following the publication of the works of Gauss and Laplace:

---

3. As cited by Harter (1974a, p.155).

> "It was for a long time more or less regarded as an axiom that statistical distributions of practically all kinds would approach the normal distribution as an ideal limiting form, if only we could dispose of a sufficiently large number of sufficiently accurate observations. The deviation of any random variable from its mean was regarded as an 'error', subject to the 'law of errors' expressed by the normal distribution."

Against this background, a number of eminent scientists of the day, such as Dirichlet, Hagen, Bessel, Laurent, Edgeworth, Poincaré, Newcomb and Charlier, began to question the universal and indiscriminate use of the Gaussian law of error.[4] Also, a slow trickle of empirical studies consistently showed that observed measurement errors appear to follow distribution laws in which larger errors occur with a higher probability than predicted by the normal distribution. For example, Bessel (1818) observed three empirical error distributions exhibiting such tendencies, but considered the discrepancies from the normal law to be insignificant. Newcomb (1886) studied the empirical distribution of 684 measurements of an angle and found the frequency of occurrence of large errors to be greater than that predicted by the Gaussian law. He boldly stated (p.346) that such empirical distributions "must be the case in nearly all astronomical and physical work". For other empirical examples of measurement errors following distributions with fatter tails than the normal distribution, see Jeffreys (1948, Ch. 5.7), and references in Hampel (1973, p.88) and Hsu (1979).

Newcomb (1886), Eddington (1933) and Jeffreys (1938) also raised theoretical objections to the Gaussian law of error. They argued that

---

4. See Harter (1974a, 1974b) for a review of the controversy surrounding the Gaussian law of error. Also see Särndal (1971) for a more detailed account of the work of Charlier and the Scandinavian school at the turn of the century.

even if one accepts the hypothesis that the error of any given measurement is normally distributed, the variance of such errors will most likely differ from observation to observation. This could occur as a consequence of the data being collected either by a number of observers of differing skill or by the same observer working under varying conditions. Lack of knowledge about how the error variance changes from measurement to measurement allows one to view the variance of individual errors as a random variable. The observed distribution of such error therefore will have a probability density function of the form

$$f(x) = \int_0^\infty (2\pi\tau^2)^{-\frac{1}{2}}\exp\{-x^2/2\tau^2\}dF(\tau), \tag{2.2.1}$$

where $F(\tau)$ is a distribution function supported on $[o,\infty)$. If, for example, $F(\tau)$ is an inverted gamma distribution function, then (2.2.1) is the density function of a Student's t distribution. Members of this latter class of error distribution have been preferred as alternatives to the Gaussian law of error by Jeffreys (1948) and Anderson and Ellis (1971).

3. THE RELEVANCE OF THE NORMALITY ASSUMPTION IN AN ECONOMIC CONTEXT

The discussion so far has centred on problems in which the stochastic variability is solely due to measurement error. This thesis is concerned with the theory of statistical inference in the linear regression model

$$y = X\beta + u, \tag{2.3.1}$$

where y is an n-dimensional random vector, X is an $n \times k$ matrix of observations on k nonstochastic variables, β is a k-dimensional vector of unknown parameters and u is an n-dimensional vector of random disturbances. In the majority of economic applications, measurement error will only be one of many possible causes of stochastic variability in (2.3.1).

Haavelmo (1944) argued that random disturbance terms of econometric models can be considered to be the sum of a large number of independent, small, elementary random shocks and, therefore, will be approximately normal. This line of reasoning has become the standard justification for the normality assumption in econometrics. In this section we critically examine its validity.

Johnston (1972, pp.10-11) gives the following three reasons for including a disturbance term in the linear regression model, (2.3.1):

(i) It captures the effects on the endogenous variable of variables not included in the model.

(ii) It allows the element of arbitrariness of human decisions to be included in the model.

(iii) It takes account of any error in the measurement of the endogenous variable.

The linear regression model (2.3.1) may be viewed as a simple approximation to a more complicated and perhaps nonlinear model. This leads to a fourth rationale for the inclusion of the disturbance term:

(iv) It allows one to take account of any discrepancy between the true and the approximating models.[5]

Whether disturbance terms can always be expressed as sums of many independent random shocks, is by no means clear. It seems reasonable

---

5. (i)-(iv) are not necessarily mutually exclusive.

to assume that they can be expressed as functions of independent random shocks, but need such functions always be linear?

Assuming these functions are linear (or approximately linear), we shall now consider whether, via the Central Limit theorem, this implies the disturbances are approximately normally distributed.

The following is a statement of the Central Limit theorem due to Lindeberg (1922):

Theorem 2.1: (Lindeberg Central Limit theorem). Let $z_1, z_2, \ldots$ be mutually independent random variables with distribution functions, $F_1, F_2, \ldots$, respectively, such that

$$E(z_i) = 0,$$
$$Var(z_i) = \sigma_i^2, \qquad i=1,2,\ldots .$$

Let

$$s_n^2 = \sigma_1^2 + \ldots + \sigma_n^2$$

and assume that for each $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} \int_{|z|\geq\varepsilon s_n} z^2 dF_i(z) = 0. \tag{2.3.2}$$

Then the distribution of the normalized sum,

$$s_n^* = (z_1+\ldots+z_n)/s_n, \tag{2.3.3}$$

tends to the standard normal distribution as $n \to \infty$.

The so-called Lindeberg condition, (2.3.2), implies that for arbitrary $\varepsilon > 0$ and n sufficiently large,

$$\sigma_k/s_n \leq \varepsilon, \qquad \text{for} \quad k=1,\ldots,n. \tag{2.3.4}$$

σ_k/s_n is a measure of the variation z_k contributes to the total variation of (2.3.3). Thus (2.3.4) may be described as stating that, asymptotically, (2.3.3) is the sum of many infinitesimal individual random shocks.

Note that (2.3.2) is not a necessary condition. The problem of whether necessary and sufficient conditions exist for the convergence of normalized sums, such as (2.3.3), to the standard normal distribution, to the author's knowledge, remains unsolved. Feller (1971) has shown (2.3.2) to be necessary in the following sense:

Theorem 2.2. (Feller) If $s_n \to \infty$ and $\sigma_n/s_n \to 0$ as $n \to \infty$ then the Lindeberg condition, (2.3.2), is necessary for the convergence of (2.3.3) in distribution to a standard normal random variable.

Returning to our original problem of whether the disturbances of (2.3.1) can be regarded as being approximately normal, the first question we should ask is whether the independent random shocks, which sum to a disturbance term, satisfy the Lindeberg condition. Perhaps the most honest answer is that we cannot always be sure. One point is clear; it is likely that the degree of heterogeneity amongst the individual random shocks will be high.[6] Whether, in general, it is great enough to violate the Lindeberg condition is uncertain. Koenker and Bassett (1978) argue that:

---

6. For example, the random shock to the world economy caused by a particular housewife's decision not to buy a certain brand of soap powder is almost infinitesimal in comparison to the random shock that might be caused by arbitrary human behaviour in a decision of OPEC oil ministers to increase oil prices.

> "... it is rather puzzling that investigators, who are generally loathe to adopt informative priors about the systematic structure of their models about which theoretical considerations and past empirical experience should provide substantive evidence, should feel themselves so well informed about the unobservable constituents of their model's unobservable errors to argue that they satisfy a Lindeberg condition! A few gross errors occurring with low probability can cause serious deviations from normality: to dismiss the possibility of these occurrences almost invariably requires a leap of Gaussian faith into the realm of pure speculation."

Suppose we take such a "leap of Gaussian faith" and assume that the Lindeberg condition[7] is satisfied. Intuitively, one could argue that the speed of convergence to the standard normal law will depend on the degree of homogeneity of the independent random shocks being summed; the more homogeneous the random shocks, the faster the convergence. For grossly heterogeneous random shocks, which may often be the case in econometrics, the approximation to normality may be extremely poor because of the slowness of convergence.

Even if for individual disturbance terms, the degree of approximation to normality is good, nothing in Haavelmo's theory guarantees the homogeneity of regression disturbances. If the disturbances of (2.3.1) are normally distributed with differing variances and we are completely ignorant about how the variance changes from observation to observation, then we may view the variance of the disturbances as a random variable. In this case, the observed distribution of the disturbances will have a probability density function of the form of (2.2.1).[8]

---

7. Or some other sufficient condition for convergence to normality.

8. C.f. Anscombe's (1967) suggestion of assuming that regression disturbances follow a Student's t distribution with seven degrees of freedom.

Finally, let us consider the assumption of independent disturbances that often accompanies the normality assumption in classical regression analysis. There is no special reason why disturbances should be independent. When one considers both the dynamic and interrelated nature of a modern economy, this assumption begins to look like wishful thinking. As Lindley (1979) recently remarked, "it is hard to find things that are truly independent".

This point is well recognised by econometricians as the large volume of literature concerned with the problem of serial correlation in regression disturbances, attests. However, only for the normal distribution does zero correlation between random variables necessarily imply independence. The lack of evidence of correlation between disturbances, therefore, should not always be interpreted as an indication of independence. The convention of assuming independent regression disturbances is perhaps a poor model for the physical world.

What does the available empirical evidence tell us about the relevance of the normality assumption? There is a considerable literature on the observed probability distributions of rates of return on common stock in various countries. Mandelbrot (1963, 1967), Fama (1963, 1965), Press (1967), Praetz (1972), Blattberg and Gonedes (1974) and Praetz and Wilson (1978) all agree on one point - that such distributions have fatter tails than predicted by the normal distribution. Carlson (1975) found some empirical distributions of price expectations displayed a similar tendency when he used survey data to test whether price expectations are normally distributed. Granger and Orr (1972) summed up the situation as follows:

> "Lately, there has been a growing awareness that some economic data display distributional characteristics that are flatly inconsistent with the hypothesis of normality. Frequency functions have been observed with too much mass near the mean and in the extreme tails, and not enough mass in the intervals between one and two standard deviations (roughly) from the mean ... . This pattern of behaviour holds true of time series and cross-section data alike, and has been observed for such widely disparate variables as stock and commodity price changes; sales, employment or asset size measures of business firms; and personal incomes."

On the other hand, Ramsey and Zarembka (1971) tested the normality assumption in five alternative specifications of a U.S. aggregate production function and found they could not reject any model solely on the basis of non-normality of the disturbances. In a similar study, Huang (1973) rejected the normality assumption of one model out of five. However, the tests used in both studies were not chosen because of their sensitivity to fat-tailed alternative distributions. This, together with the fact that approximately only 50 degrees of freedom were available for testing, suggests that these findings should not be regarded as a vindication of the normality assumption.

4. CONCLUDING REMARKS

In summary, it is clear that there are good theoretical reasons for questioning the standard assumption of normal disturbances in the linear regression model set in an economic context. The available empirical evidence tends to support these reasons. Non-normality of regression disturbances has a number of implications for the properties of least squares estimators and standard hypothesis testing procedures.

The main response to this growing disenchantment with the normality assumption has been the development of a number of robust alternatives to the least squares estimator. This literature is reviewed by Huber (1972, 1973), Hampel (1973) and Harter (1974-1975). The widespread use of such estimation techniques has perhaps been hampered slightly by a lack of suitable statistical procedures for conducting associated tests of hypotheses.[9]

In the subsequent three chapters of this thesis, a different approach to the problem of non-normal disturbances is considered. We investigate the implications for the classical theory of regression analysis of replacing the normality assumption with the broader assumption that the regression disturbances follow an elliptically symmetric distribution.

This class of distributions was chosen for two reasons. First, a large subclass of elliptically symmetric distributions have univariate marginal distributions with probability density functions of the form of (2.2.1). This class of disturbance distributions, therefore, answers one of the theoretical objections to the normality assumption raised in the previous section. Secondly, recent work by Zellner (1976),[10] Kariya and Eaton (1977) and Kariya (1977) indicates that elliptically symmetric distributions possess properties which make them analytically tractable. These studies found that the Type I errors of the standard F tests and

9. Such statistical procedures are beginning to be developed. For example, see Bickel (1978).

10. Zellner principally considered spherically symmetric Student's t distributions which are a special case of elliptically symmetric distributions.

Student's t tests of regression parameter values and the Durbin-Watson test, are invariant to a widening of the normality assumption to include appropriate elliptically symmetric disturbance distributions. Kariya showed that for a particular subclass of elliptically symmetric disturbance distributions, the Durbin-Watson test is an approximately uniformly most powerful invariant test for first-order autoregressive disturbances whenever the column space of the design matrix is spanned by k of the characteristic vectors of the first difference matrix, where k is the number of regressors. Clearly these results encourage further research.

Although the assumption of elliptically symmetric disturbances answers at least one of the criticisms raised against the normality assumption, we are by no means certain that it provides the complete answer. Any meaningful assessment of its usefulness can only come after a thorough investigation of the properties of the linear regression model with elliptically symmetric disturbances. Further judgement of the merits of this broadening of the normality assumption is therefore left until Chapter 9.

CHAPTER 3

SPHERICALLY SYMMETRIC AND ELLIPTICALLY SYMMETRIC DISTRIBUTIONS: A REVIEW OF THE LITERATURE

1. INTRODUCTION

The purpose of this chapter is to introduce the reader to spherically symmetric and elliptically symmetric distributions. Definitions and examples are given in the next section and this is followed in subsequent sections by a review of the literature dealing with their various properties. We shall make use of many of these properties in later chapters. For this reason, the final section contains a convenient summary of the more useful properties that have been derived in the literature.

2. DEFINITIONS AND EXAMPLES

An n-dimensional random vector, x, has a spherically symmetric distribution[1] if its distribution laws are invariant to rotations about the origin, i.e. to orthogonal transformations. Clearly the distribution laws of such random vectors will be independent of direction from the origin and therefore will be a function only of distance from the origin, namely $r = (x'x)^{\frac{1}{2}}$. In particular this means that if x has a joint probability density function, it will be of the form

(3.2.1) $$f(x) = \phi(x'x)$$

---

1. Also known as spherical or radial distributions in the literature.

with respect to the Lebesgue measure on R^n, where

$$(3.2.2) \qquad \phi : [0,\infty) \to [0,\infty)$$

and

$$(3.2.3) \qquad \int_0^\infty r^{n-1}\phi(r^2)\,dr = \tfrac{1}{2}\Gamma(n/2)\pi^{-n/2}$$

Two alternative definitions of an n-dimensional spherically symmetric random vector used in the literature are: (i) a random vector which has a joint density function of the form of (3.2.1) and (ii) a random vector with a joint characteristic function of the form

$$(3.2.4) \qquad \psi(t) = \Psi(t't)$$

where Ψ is some function on $[0,\infty)$ and t is an n-dimensional vector. The latter is equivalent to the definition given above while the former is more restrictive as it defines only those spherically symmetric distributions for which joint density functions exist.

Spherically symmetric distributions are members of a wider class of distributions which we shall call elliptically symmetric.[2] An n-dimensional random vector z has an elliptically symmetric distribution if its distribution laws are invariant to orthogonal transformations in the n-dimensional Euclidean geometry with the inner product

$$(u,v) = [u'\Sigma^{-1}v]^{\frac{1}{2}}$$

---

2. These distributions have also been called elliptical [Chu (1973)], spherically invariant [Vershik (1964)], elliptically contoured [DasGupta et al. (1972)] distributions and even occasionally spherically symmetric distributions [Goldman (1976)].

where u and v are n-dimensional vectors and Σ is a given, n x n, positive definite matrix. Σ is termed the characteristic matrix.[3]

Throughout this thesis the notation $E(n,\Sigma)$ will be used to denote an n-dimensional elliptically symmetric distribution with characteristic matrix Σ . Obviously a spherically symmetric distribution is a special case of an $E(n,\Sigma)$ distribution where $\Sigma = \sigma^2 I_n$ and σ^2 is any positive scalar. Thus $E(n,I_n)$ denotes an n-dimensional spherically symmetric distribution. Further, the notation

$$z \sim E_o(n,\Sigma)$$

will be used to denote that z is an $E(n,\Sigma)$ random vector such that

$$\Pr(z=0) = 0 \text{ .}$$

Clearly, if z is an $E(n,\Sigma)$ random vector and S is any n x n, nonsingular matrix such that

$$S'S = \Sigma^{-1} \text{ ,}$$

then

(3.2.5) $$x = Sz$$

---

3. For any given elliptically symmetric distribution the characteristic matrix is uniquely determined up to a scalar factor. Wise and Gallagher (1978) widened the definition of elliptically symmetric distributions by allowing Σ to be non-negative definite thus including singular random vectors. (A random vector is singular if one of its components can be expressed as a linear combination of the other components.)

has a spherically symmetric distribution. On the other hand, if x is an $E(n,I_n)$ random vector with a joint density function of the form of (3.2.1) and if A is an $n \times n$ non-singluar matrix then

$$z = Ax \tag{3.2.6}$$

has a joint density function of the form

$$g(z) = |A^{-1}|\phi(z'(A^{-1})'A^{-1}z)$$

$$= |\Sigma|^{-\frac{1}{2}}\phi(z'\Sigma^{-1}z) \tag{3.2.7}$$

with respect to the Lebesgue measure on R^n where $\Sigma = AA'$. In view of (3.2.5) and (3.2.6), if z is an $E(n,\Sigma)$ random vector with a joint density function, then this function will be of the form of (3.2.7) where ϕ is such that (3.2.2) and (3.2.3) hold.

As in the spherically symmetric case, two alternative definitions of an $E(n,\Sigma)$ random vector have been used in the literature. They are: (i) a random vector which has a joint density function of the form of (3.2.7) and (ii) a random vector with a joint characteristic function of the form

$$\psi(t) = \Psi(t'\Sigma t) \tag{3.2.8}$$

where Ψ is some function on $[0,\infty)$ and t is an n-dimensional vector. Again the latter definition is equivalent to our definition while the former is more restrictive, defining only those elliptically symmetric distributions for which joint density functions exist.[4]

---

4. For a proof of the equivalence of (i) and (ii) in the case when a joint density function of the $E(n,\Sigma)$ random vector is assumed to exist see Blake and Thomas (1968) or Goldman (1976).

Any n-dimensional random vector with a joint density function of the form of (3.2.7) for some $n \times n$ positive definite matrix Σ is clearly elliptically symmetric. The class of elliptically symmetric distributions includes the multivariate normal distribution, $N(o,\sigma^2\Sigma)$, and the multivariate Student's t distribution with a joint density function of the form

$$(3.2.9)\qquad g(z) = p(\nu_o)(\sigma^2)^{-n/2}|\Sigma|^{-\frac{1}{2}}\{\nu_o + z'\Sigma^{-1}z/\sigma^2\}^{-(n+\nu_o)/2}\ ,$$

$$\sigma,\nu_o>o,\ -\infty<z_i<\infty,\ i=1,\ldots,n,$$

where

$$p(\nu_o) = \nu_o^{\nu_o/2}\Gamma[(\nu_o+n)/2]/\pi^{n/2}\Gamma(\nu_o/2)\ .$$

For $\nu_o > 2$, and $\Sigma = I_n$, the elements of z are uncorrelated but not independent. When $\nu_o = 1$, (3.2.9) is a multivariate Cauchy density function for which no moments exist.

An important class of elliptically symmetric distributions are those whose joint density functions can be expressed as mixtures of multivariate normal densities, i.e. as the Lebesgue-Stieltjes integral,

$$(3.2.10)\qquad g(z) = \int_o^\infty (2\pi\tau^2)^{-n/2}|\Sigma|^{-\frac{1}{2}}\exp\{-z'\Sigma^{-1}z/2\tau^2\}dF(\tau)$$

where $F(\tau)$ is any distribution function supported on $[o,\infty)$. In fact for every elliptically symmetric density function of the form (3.2.7),

$$h(z) = \int_o^\infty \tau^{-n}|\Sigma|^{-\frac{1}{2}}\phi(z'\Sigma^{-1}z/\tau^2)\,dF(\tau)$$

is also an $E(n,\Sigma)$ joint density function. Further examples of elliptically symmetric distributions have been given by Lord (1954), McGraw and Wagner (1968), Chu (1973) and Goldman (1976).

The literature on elliptically symmetric distributions can be classified into three groups: (i) papers investigating the properties of $E(n,\Sigma)$ random vectors, (ii) those that deal with the class of spherically invariant stochastic processes; that is stochastic processes whose n-dimensional marginal distributions are $E(n,\Sigma)$ and (iii) papers concerned with stochastic matrices whose rows (or columns) are $E(n,I_n)$; such matrices being labelled spherical random matrices. Contributions in each of these three branches of the literature are surveyed chronologically in the following three sections.

3. ELLIPTICALLY SYMMETRIC RANDOM VECTORS

Early work on spherically symmetric distributions was done by Maxwell (1860) who showed that independence of the components of an $E(n,I_n)$ random vector implied normality.[5]

Lord (1954) derived the joint characteristic function, $\psi(t)$, of an $E(n,I_n)$ random vector, x, with joint density function (3.2.1); i.e.,

$$\psi(t) = \int_{R^n} \exp\{ix't\}\phi(x'x)dx$$

---

5. Bartlett (1934) confirmed this result. Its implications for the significance of the majority of results reported in Chapter 4 and 5 is discussed in Chapter 9.

where t is an n-dimensional vector. He found it to be the Hankel transform:

$$(3.3.1) \qquad \psi(t) = (2\pi)^{n/2}\|t\|^{-n/2+1}\int_0^\infty r^{n/2}J_{n/2-1}(r\|t\|)\phi(r^2)\,dr$$

where $\|t\| = (t't)^{\frac{1}{2}}$ and J_m is the Bessel function of the first kind of m^{th} order. Clearly $\psi(t)$ is a function only of $\|t\|$, hence confirming that (3.2.4) holds when x has a joint density function.

Lord used this form of the characteristic function to show that projections of spherically symmetric distributions onto spaces of fewer dimensions are spherically symmetric and that when the characteristic functions of distributions so related are expressed in the form of (3.2.4), Ψ takes a common form. He also demonstrated that the sum of independent, not necessarily identically distributed, $E(n,I_n)$ random vectors is $E(n,I_n)$.

The joint characteristic function of an $E(n,\Sigma)$ random vector, z, with joint density of the form (3.2.7) is

$$(3.3.2) \qquad \int_{R^n} \exp\{iz't\}|\Sigma|^{-\frac{1}{2}}\phi(z'\Sigma^{-1}z)\,dz \ .$$

Note that by applying the transformation (3.2.5) and following Lord's proof for the spherically symmetric case, (3.3.2) can be shown to be the Hankel transform (3.3.1) with $\|t\| = (t'\Sigma t)^{\frac{1}{2}}$. This allows us to confirm that the joint characteristic function of an $E(n,\Sigma)$ random vector with joint density function (3.2.7) is of the form (3.2.8).

Box and Hunter (1957) found expressions for the moments of a spherically symmetric random vector while Kingman (1963) studied the one dimensional random walk generated by distance from the origin in R^n space when independent $E(n,I_n)$ random vectors are summed.

A number of significant contributions to the theory of elliptically symmetric distributions were made by Kelker (1970). He noted that a sufficient condition for the existence of the k^{th} order moment of an $E(n,\Sigma)$ random vector with joint density of the form (3.2.7) is that

$$\int_0^\infty r^{k+n-1}\phi(r^2)\,dr < \infty \qquad (3.3.3)$$

holds.[6] If (3.3.3) is satisfied for $k = 1$ then the mean of the $E(n,\Sigma)$ random vector, z, is 0 while if it holds for $k = 2$, the covariance matrix of z is $\sigma^2\Sigma$ where σ^2 is a positive scalar, independent of Σ. In the case $\Sigma = I_n$, σ^2 is the variance of each of the univariate marginal distributions.

Suppose z is an $E(n,\Sigma)$ random vector and that z and Σ are partitioned as

$$z = \begin{bmatrix} u \\ v \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where u and v have $n-q$ and q components respectively and Σ_{11} and Σ_{22} are $(n-q) \times (n-q)$ and $q \times q$ respectively. Kelker found the marginal density of v to be of the form

---

6. Note that if (3.3.3) holds for $k = k_1$ it also holds for $0 \leqslant k < k_1$.

(3.3.4) $$f(v) = |\Sigma_{22}|^{-\frac{1}{2}}\phi_q(v'\Sigma_{22}^{-1}v)$$

for some function ϕ_q where ϕ_q is determined only by the form of ϕ and by q. Thus all marginal density functions of dimension q, $q < n$, have the same functional form. Further, he found that when $E(u|v)$ exists,

(3.3.5) $$E(u|v) = \Sigma_{12}\Sigma_{22}^{-1}v \ .$$

In particular, this implies that if $\begin{bmatrix} u \\ v \end{bmatrix}$ is an $E(n,I_n)$ random vector,

(3.3.6) $$E(u|v) = 0 \ .$$

He also demonstrated that assuming the existence of a joint density function is not particularly restrictive. This was done by showing that all marginal distributions of an $E_o(n,\Sigma)$ distribution have probability density functions. Therefore a sufficient condition for an $E(n,\Sigma)$ distribution to have a joint density function is that it be a marginal distribution of an $E_o(n+1,\bar{\Sigma})$ distribution where $\bar{\Sigma}$ is any $(n+1) \times (n+1)$ positive definite matrix. Kelker generalized Maxwell's early result by showing that if Σ is a diagonal matrix, then independence of the components of an $E(n,\Sigma)$ random vector, z, implies z has a $N(0,\sigma^2\Sigma)$ distribution. In addition he proved[7] that when $\{x_n\}_{n=1}^{\infty}$ is a countable family of random variables, the joint distribution of any k of these variables is $E(k,I_k)$ for each k, k=2, 3, ... if and only if there is a non-negative random variable, τ, such that conditional on the value taken by τ, the

---

7. For two alternative proofs of this result see Kingman (1972).

x_n's are independent normal variables with mean zero and variance τ, i.e., the joint density has the form (3.2.10) with $n = k$ and $\Sigma = I_k$. A further property he reported is that if x is an $E(n, I_n)$ random vector with a joint density function, then the random variable

$$\sum_{i=1}^{n-q} x_i^2 \Big/ \sum_{i=1}^{n} x_i^2$$

has a beta distribution with parameters $(n-q)/2,\ q/2$.

Chu (1973) showed that if z is an $E(n,\Sigma)$ random vector with joint density function (3.2.7), there exists a scalar function $w(s)$ defined on (o,∞) such that

$$\int_o^\infty w(s)\,ds = 1$$

and (3.2.7) can be written as

$$g(z) = \int_o^\infty w(s)\,\Phi_N(z,\Sigma/s)\,ds \tag{3.3.7}$$

where $\Phi_N(z,\Sigma/s) = s^{n/2}(2\pi)^{-n/2}|\Sigma|^{-\frac{1}{2}}\exp\{-sz'\Sigma^{-1}z/2\}$ is the $N(o,\Sigma/s)$ joint density function. Note that $w(s)$ is not necessarily a probability density function since it may take negative values. In proving this result, Chu required $L^{-1}[\phi(r)]$ to exist,[8] where L^{-1} is the inverse Laplace transform operator. Hence (3.3.7) should not be treated as necessarily holding for all $E(n,\Sigma)$ random vectors with joint density functions.

---

8. $\phi(r)$ differentiable for sufficiently large r and $r^k\phi(r) \to o$ as $r \to \infty$ for some $k > 1$ are sufficient conditions for the existence of $L^{-1}[\phi(r)]$.

Chu used (3.3.7) to demonstrate a number of properties of $E(n,\Sigma)$ random vectors. For example, if $m(z)$ is any Borel measurable function of z for which $E[m(z)]$ exists then

$$E[m(z)] = \int_0^\infty w(s) E_{N_s}[m(z)]ds$$

where $E_{N_s}[m(z)]$ is the expectation of $m(z)$ assuming z to be $N(0,\Sigma/s)$. Also if

$$(3.3.8) \qquad u = Az$$

where A is an $m \times n$ matrix of rank $m \leqslant n$ then the joint density function of u is

$$h(u) = \int_0^\infty w(s) \Phi_N(u, A\Sigma A'/s) ds$$

$$(3.3.9) \qquad = |A\Sigma A'|^{-\frac{1}{2}} \phi_m(u'(A\Sigma A')^{-1}u) \; .$$

Note that $h(u)$ has the same weighting function as $g(z)$ in (3.3.7). Thus the change in form of ϕ is caused by the change in the functional form of the normal density as the dimension changes from n to m. (3.3.9) holds for transformations of any $E(n,\Sigma)$ random vector with a joint density function and can also be obtained using (3.3.4). Further, it is clear from the definition of elliptically symmetric distributions that if z is an $E(n,\Sigma)$ random vector, not necessarily with a joint density function, then u defined by (3.3.8) will be an $E(m,A\Sigma A')$ random vector.

Goldman (1976) noted that the sum of two independent elliptically symmetric random vectors with different characteristic matrices is not in general elliptically symmetric, notable exceptions being when the two random vectors are normally distributed or when the two characteristic matrices differ only by a scalar factor. He also showed that if x is an $E(n,I_n)$ random vector with joint density function (3.2.1), there exists a unique set of random variables:

$$r \in [o,\infty),$$
$$\theta_k \in [o,\pi], \qquad k=1, \ldots, n-2,$$
$$\theta_{n-1} \in [o,2\pi)$$

for which

$$(3.3.10) \quad \begin{cases} x_1 = r\cos\theta_1, \\ x_j = r(\prod_{k=1}^{j-1}\sin\theta_k)\cos\theta_j \qquad 2\leqslant j\leqslant n-1, \\ x_n = r\prod_{k=1}^{n-1}\sin\theta_k \end{cases}$$

Furthermore, r, θ_1, ..., θ_{n-1} are independent and have respective density functions

$$(3.3.11) \quad \begin{cases} p_r(r) = 2\pi^{n/2}[\Gamma(n/2)]^{-1}r^{n-1}\phi(r^2), \qquad r \in [o,\infty),^{9} \\ p_{\theta_k}(\theta_k) = \Gamma[(n-k+1)/2]\pi^{-\frac{1}{2}}[\Gamma[(n-k)/2]]^{-1}\sin^{n-1-k}\theta_k, \\ \qquad\qquad \theta_k \in [o,\pi],\ k=1, \ldots, n-2, \\ p_{\theta_{n-1}}(\theta_{n-1}) = {}^{1}/(2\pi), \qquad \theta_{n-1} \in [o,2\pi) . \end{cases}$$

Consequently, if r, θ_1, ..., θ_{n-1} are independent random variables with density functions given by (3.3.11) then the random vector defined by (3.3.10) is $E(n,I_n)$.

A further consequence is that if x is an $E(n,I_n)$ random vector then its density function (if it exists) can be written as

9. Note that $\int_o^\infty p_r(r)dr = 1$ implies (3.2.3).

$$f(x) = \phi(x'x)$$
$$= (2\pi^{n/2})^{-1}\Gamma(n/2)p_{\|x\|}[(x'x)^{\frac{1}{2}}]/(x'x)^{(n-1)/2}$$

where $p_{\|x\|}[\cdot]$ is the marginal density of distance from the origin. This can be extended to the case where z is an $E(n,\Sigma)$ random vector by making use of the transformation (3.2.5) so that

$$g(z) = |\Sigma|^{-\frac{1}{2}}\phi(z'\Sigma^{-1}z)$$
$$= (2\pi^{n/2})^{-1}\Gamma(n/2)|\Sigma|^{-\frac{1}{2}}p_{\|Sz\|}[(z'\Sigma^{-1}z)^{\frac{1}{2}}]/(z'\Sigma^{-1}z)^{(n-1)/2}$$

where S is any $n \times n$, nonsingular matrix such that $S'S = \Sigma^{-1}$.

Kariya and Eaton (1977) considered the distributions of $x/(x'x)^{\frac{1}{2}}$, $a'x/(x'x)^{\frac{1}{2}}$ and $x'Ax/(x'x)^{\frac{1}{2}}$ where x is an $E_o(n,I_n)$ random vector, a is an n-dimensional vector such that $a'a = 1$ and A is an $n \times n$ symmetric matrix. $x/(x'x)^{\frac{1}{2}}$ was shown to be uniformly distributed on the n-dimensional, unit, hypersphere,

$$C_n = \{x | x \in R^n,\ x'x = 1\} .$$

$(n-1)^{\frac{1}{2}}w/(1-w^2)^{\frac{1}{2}}$, where $w = a'x/(x'x)^{\frac{1}{2}}$, was found to have a Student's t distribution with $n - 1$ degrees of freedom. $x'Ax/(x'x)^{\frac{1}{2}}$ was shown to have the same distribution as $\sum_{j=1}^{n} d_j y_j$ where d_j, $j=1, \ldots, n$, are the latent roots of A and $y_1, \ldots, y_{n-1}$ have the joint density function,

$$p(y_1, \ldots, y_{n-1}) = \Gamma(n/2)[\Gamma(\tfrac{1}{2})]^{-n}[\prod_{i=1}^{n-1} y_i^{-\frac{1}{2}}](1-\sum_{i=1}^{n-1} y_i)^{-\frac{1}{2}}$$

with $y_n = 1 - \sum_{i=1}^{n-1} y_i$. In the special case when $A^2 = A$ and $\text{rank}(A) = k$, $x'Ax/(x'x)^{\frac{1}{2}}$ has a beta distribution with parameters $k/2,\ (n-k)/2$.

Others who have made contributions to the theory of elliptically symmetric distributions include Baldessari (1965), Ahmad (1972), Das Gupta et al. (1972), Goldman (1972), Strawderman (1974), Wolfe (1975), Ghosh and Pollak (1975), Crawford (1977) and Kariya (1977).

4. SPHERICALLY INVARIANT STOCHASTIC PROCESSES

A stochastic process may be regarded as an indexed set of random variables, $\{x_t, t \in T\}$, defined on a common probability space $(\Omega, \mathcal{F}, P)$. A spherically invariant stochastic process is a stochastic process for which the joint distribution of each finite arbitrary collection of the x_t, $t \in T$, is elliptically symmetric. This class of stochastic process was first considered by Vershik (1964) who restricted his attention to zero-mean, square integrable stochastic processes. He defined a process to be spherically invariant if in the linear manifold generated by arbitrary samples from the process, all random variables having the same variance have the same distribution function. Blake and Thomas (1968) established that this definition is equivalent to the former definition in the restricted class of processes studied by Vershik.

Using the concept of semi-independency, Vershik found that his class of spherically invariant stochastic processes is uniquely characterized by the fact that every mean-square estimation problem on the linear manifold generated by such a process has a linear solution.[10] Two random variables, x_1 and x_2, are semi-independent

---

10. The best mean-square estimator of x_{t_o}, $t_o \in T$, given $x_{t_1}, \ldots, x_{t_n}$; $t_1, \ldots, t_n \in T$, is the conditional expectation $E[x_{t_o} | x_{t_1}, \ldots, x_{t_n}]$. Every mean-square estimation problem has a linear solution if $E[x_{t_o} | x_{t_1}, \ldots, x_{t_n}]$ is a linear function for arbitrary n and any $\{t_o, t_1, \ldots, t_n\} \subset T$.

if

$$E[x_i | x_j] = E[x_i] \qquad i,j=1,2,i\neq j,$$

or equivalently if each random variable is uncorrelated with any arbitrary function of the other.[11] Clearly from (3.3.6) the components of an $E(n,I_n)$ random vector whose second order moments exist are semi-independent. Vershik also demonstrated that his class of spherically invariant processes is precisely that class of processes for which the concepts of wide and narrow stationarity are equivalent.

Yao (1973), who unlike Vershik did not restrict attention to square integrable processes, proved that a process is spherically invariant if and only if it is equivalent to a stochastic process of the form $\{\alpha y_t,\ t \in T\}$ where α is a non-negative random variable and $\{y_t,\ t \in T\}$ is a zero-mean Gaussian stochastic process independent of α. Therefore any n-dimensional sample from a spherically invariant stochastic process has a joint density function[12] of the form (3.2.10) where $F(\cdot)$ is the distribution function of α. On the other hand, for a distribution to belong to a finite dimensional sample from a stochastic process it must satisfy the Kolmogorov consistency conditions[13] so that any $E(n,\Sigma)$

---

11. The concept of semi-independence lies between that of independence and being uncorrelated, in the sense that when second-order moments are assumed to exist, independence implies semi-independence and semi-independence implies zero correlation.

12. In view of Kelker's (1970) findings, joint density functions will exist for all such n-dimensional samples provided $Pr(x_t=0) = 0$ for all $t \in T$.

13. See Cramer and Leadbetter (1967, p.30).

distribution is not necessarily the distribution of an n-dimensional sample from a spherically invariant process and hence need not have a joint density function of the form (3.2.10). Yao also noted that if $\{x_n, n \geq 1\}$ is a sequence of jointly spherically symmetric random variables with finite second order moments, then $\{z_n, n \geq 1\}$, where $z_n = \sum_{i=1}^{n} x_i$, is a martingale.

Yao's characterisation of spherically invariant stochastic processes was recently extended by Wise and Gallagher (1978) to include singular[14] spherically invariant stochastic processes and to allow α to be any real valued random variable independent of $\{y_t, t \in T\}$. Contributions to the theory of spherically invariant stochastic processes have also been made by Picinbono (1970) and Gualtierotti (1974, 1976).

5. SPHERICAL MATRIX DISTRIBUTIONS

Spherical matrix distributions have been studied by Maxwell (1860), Dempster (1969) and Dawid (1977). They are of interest because they arise naturally in certain applications, a prominant example being quantum theory [see Mehta (1967)]. Dawid defines an $n \times p$ random matrix, Y, to be left-spherical if its joint distribution is invariant to transformations of the form,

$$Y \rightarrow UY$$

and right-spherical if its joint distribution is invariant to

---

14. A stochastic process is singular if one random variable in the process can be expressed as a finite linear combination of the other random variables in the process.

transformations of the form

$$Y \to YV$$

where U and V are nonstochastic orthogonal matrices of dimensions (n x n) and (p x p) respectively. Clearly, the columns of a left-spherical random matrix and the rows of a right-spherical random matrix are spherically symmetric.

Dawid investigated problems of inference about the parameters of a multivariate linear model in which the usual assumption of normality for the disturbances is replaced by the weaker assumption that they follow an appropriate spherical matrix distribution. He found that inference about parameter values and associated confidence regions are identical to whatever is appropriate under normality.

6. A SUMMARY OF THE IMPORTANT PROPERTIES OF ELLIPTICALLY SYMMETRIC DISTRIBUTIONS

Throughout this section, z denotes an $E(n,\Sigma)$ random vector.

(I) If the joint density function of z exists, then it is of the form

$$g(z) = |\Sigma|^{-\frac{1}{2}}\phi(z'\Sigma^{-1}z) \tag{3.6.1}$$

with respect to the Lebesgue measure on R^n where

$$\phi : [0,\infty) \to [0,\infty)$$

and

$$\int_0^\infty r^{n-1}\phi(r^2)dr = \tfrac{1}{2}\Gamma(n/2)\pi^{-n/2}$$

(II) z is an $E(n,\Sigma)$ random vector if and only if it has a characteristic function of the form

$$\psi(t) = \Psi(t'\Sigma t)$$

where t is an n-dimensional vector and Ψ is a function on $[o,\infty)$ independent of Σ.

(III) If A is an $m \times n$ matrix of rank m where $m \leq n$, then Az is $E(m,A\Sigma A')$. In particular, all marginal distributions of any elliptically symmetric distribution are elliptically symmetric. If z has a joint density function of the form of (3.6.1), then $w = Az$ has a joint density function of the form

$$f(w) = |A\Sigma A'|^{-\frac{1}{2}}\phi_m(w'(A\Sigma A')^{-1}w)$$

where ϕ_m is determined only by the form of ϕ and by m and is such that

$$\phi_m\colon [o,\infty) \to [o,\infty)$$

and

$$\int_o^\infty r^{m-1}\phi_m(r^2)\,dr = \tfrac{1}{2}\Gamma(m/2)\pi^{-m/2}$$

(IV) If Σ is a diagonal matrix, independence of the components of z implies z has a multivariate normal distribution.

(V) If Σ is diagonal and if the second order moments of z exist, the components of z are semi-independent. Further, if z is partitioned as

$$z = \begin{bmatrix} u \\ v \end{bmatrix}$$

where u is $(n-q)$ and v is q-dimensional, $o < q < n$, then

$$E[u|v] = 0$$

(VI) If the first order moments of z exist

$$E[z] = 0$$

while if the second order moments exist, z has convariance matrix

$$E[zz'] = \sigma^2\Sigma$$

where σ^2 is a positive scalar.

(VII) The sum of indpendent, not necessarily identically distributed elliptically symmetric vectors with characteristic matrices which differ only by scalar factors, is elliptically symmetric.

(VIII) A sufficient condition for z to have a joint density function is that it be a marginal distribution of an (n+1)-dimensional elliptically symmetric distribution which does not have an atom of weight at the origin.

(XI) A stochastic process is spherically invariant if and only if it is equivalent to a stochastic process of the form $\{\alpha y_t, t \in T\}$ where α is a non-negative random variable, and $\{y_t, t \in T\}$ is a zero-mean Gaussian stochastic process independent of α. The characteristic function of an n-dimensional sample from a spherically invariant stochastic process is of the form

$$\psi_n(s_n) = \int_o^\infty \exp(-\nu^2 s_n' \Sigma_n s_n/2)\,dF(\nu)$$

where s_n is an n-dimensional vector, Σ_n is an $n \times n$ positive definite matrix and $F(\cdot)$ is the probability distribution function of α . The sample's probability density function (if it exists) is of the form

$$h(z) = \int_o^\infty (2\pi\tau^2)^{-n/2} |\Sigma_n|^{-\frac{1}{2}} \exp(-z'\Sigma_n^{-1} z/2\tau^2)\,dF(\tau)$$

For any given spherically invariant process, $F(\cdot)$ is common to all finite samples.

(X) If x is an $E_o(n,I_n)$ random vector then

$$x/(x'x)^{\frac{1}{2}}$$

is uniformly distributed on the n-dimensional unit hypersphere:

$$C_n = \{x | x \in R^n,\ x'x = 1\}$$

(XI) If x is an $E(n,I_n)$ random vector with a joint density function of the form

$$f(x) = \phi(x'x),$$

there exists a unique set of random variables,

$$r \in [o,\infty) ,$$

$$\Theta_k \in [o,\pi] , \qquad k=1,\ldots,n-2,$$

$$\Theta_{n-1} \in [o,2\pi),$$

for which

$$x_1 = r\cos\Theta_1,$$

$$x_j = r\left(\prod_{k=1}^{j-1}\sin\Theta_k\right)\cos\Theta_j, \qquad j=2,\ldots,n-1,$$

$$x_n = r\prod_{k=1}^{n-1}\sin\Theta_k$$

Furthermore, r, Θ_1, ..., Θ_{n-1} are independent and have respective density functions:

$$p_r(r) = 2\pi^{n/2}[\Gamma(n/2)]^{-1}r^{n-1}\phi(r^2), \qquad r \in [o,\infty)$$

$$p_{\Theta_k}(\Theta_k) = \Gamma[(n-k+1)/2]\pi^{-\frac{1}{2}}[\Gamma[(n-k)/2]]^{-1}\sin^{n-1-k}\Theta_k, \quad \Theta_k \in [0,\pi]$$

$$k=1,\ldots,n-2$$

$$p_{\Theta_{n-1}}(\Theta_{n-1}) = {}^1/(2\pi) \qquad \Theta_{n-1} \in [o,2\pi)$$

CHAPTER 4

STATISTICAL PROPERTIES OF LEAST SQUARES REGRESSION ESTIMATORS WHEN DISTURBANCES ARE ELLIPTICALLY SYMMETRIC

1. INTRODUCTION

The purpose of this chapter is to investigate some of the more important statistical properties of least squares estimators of β in the usual linear regression model,

$$y = X\beta + u,$$

when the disturbance vector, u, is assumed to take an elliptically symmetric distribution.

Section 2 opens with a review of the literature on weak consistency of the OLS (Ordinary Least Squares) estimator of β. Then necessary and sufficient conditions are found for the weak consistency of linear "unbiased" estimators of β assuming elliptically symmetric disturbances. We define such an estimator to be "unbiased" if it is unbiased in the classical sense when the disturbances have finite first order moments. The remainder of Section 2 is devoted to demonstrating the versatility of this result by applying it to the OLS, GLS (Generalized Least Squares), instrumental variable and restricted least squares estimators of β.

Strong consistency of least squares estimators of β is the subject of Section 3. First, recent literature on strong consistency of the OLS estimator is reviewed and then, before our attention is confined to the case of elliptically symmetric disturbances, sufficient conditions are found for the strong consistency of the OLS estimator under the classical assumption of uncorrelated disturbances. This result allows us to obtain sufficient conditions for the strong consistency of the GLS estimator when the disturbance covariance matrix is known up to a scalar value; this being an important problem which seems to have been ignored in the literature. Finally, necessary and sufficient conditions are determined for the OLS estimator to be strongly consistent assuming the regression disturbances are elliptically symmetric with a diagonal covariance matrix and finite second order moments.

The asymptotic distributions of the OLS and GLS estimators are discussed briefly in Section 4. A powerful optimality property of the GLS estimator when the regression disturbances are $E(n,\Sigma)$ is proven in Section 5. Without any assumption being made about the existence of moments of u, the GLS estimator is shown to be better than any other linear "unbiased" estimator of β. The optimality criterion used in this result is one which is intuitively desirable as well as being more stringent than that used in the Gauss-Markov theorem.

In Section 6, the GLS estimator is shown to be the maximum likelihood estimator of β when u is $E(n,\Sigma)$ with a joint density function of the form (3.2.7) such that ϕ is a non-increasing function on $[o,\infty)$. The maximum likelihood estimator of σ^2 when u has finite covariance matrix $\sigma^2\Sigma$, is also investigated. Some concluding remarks are made in Section 7.

2. WEAK CONSISTENCY OF LEAST SQUARES REGRESSION ESTIMATORS

In this and the two subsequent sections, we shall be concerned with the model

$$(4.2.1) \qquad y_t = \beta' x_t + u_t, \qquad t=1,2,\ldots,$$

where β is a k-dimensional vector of unknown parameters, x_t is a k-dimensional nonstochastic vector and u_t is a scalar disturbance term. The first n equations of (4.2.1) can be written in matrix notation as

$$(4.2.2) \qquad y(n) = X(n)\beta + u(n),$$

where $X(n)$ is an $n \times k$ matrix which is assumed to be of rank k when $n > k$, while $y(n)$ and $u(n)$ are n-dimensional vectors.

When the only assumptions made about the joint distribution of the disturbance terms, u_t, $t = 1, 2, \ldots,$ are

$$(4.2.3) \qquad \begin{cases} E(u_t) = 0, & \\ E(u_t^2) = \sigma_t^2, & t=1,2,\ldots, \\ \sup\limits_{t\geq 1} \sigma_t^2 < \infty & \text{and} \\ E(u_t u_s) = 0, & s,t=1,2,\ldots,\ s\neq t, \end{cases}$$

Eicker (1963) proved that a sufficient condition for the OLS estimator of β,

$$(4.2.4) \qquad b(n) = (X'(n)X(n))^{-1}X'(n)y(n),$$

to be weakly consistent is that

$$(4.2.5) \qquad (X'(n)X(n))^{-1} \to 0 \quad \text{as} \quad n \to \infty.$$

He also showed that if u_t, $t = 1, 2, \ldots$ are independently, identically $N(0,\sigma^2)$, (4.2.5) is a necessary condition for weak consistency of $b(n)$.

Drygas (1971, 1976) considered the case of the disturbance vector, $u(n)$, of (4.2.2) having mean zero and arbitrary covariance matrix,

$$E(u(n)u'(n)) = \Sigma(n).$$

Let $\lambda_{max}(\Sigma(n))$ and $\lambda_{min}(\Sigma(n))$ denote the largest and smallest eigenvalues of $\Sigma(n)$ respectively for $n = 1, 2, \ldots$. Drygas extended Eicker's results to regressions with correlated disturbances when he proved that if

$$(4.2.6) \qquad \sup_{n\geq 1}\{\lambda_{max}(\Sigma(n))\} < \infty$$

and

$$(4.2.7) \qquad \inf_{n\geq 1}\{\lambda_{min}(\Sigma(n))\} > 0,$$

the OLS estimator of β is weakly consistent if and only if (4.2.5) holds. Drygas also extended Eicker's results to the case when $X(n)$ is not necessarily of full rank for $n = k, k + 1, \ldots$. He proved that if the linear combination $a'\beta$, where a is a k-dimensional vector, is estimable[1] and if both (4.2.6) and (4.2.7) hold then the

---

1. $a'\beta$ is estimable if a is a linear combination of the rows of $X(n)$ for some positive integer n.

least squares estimator of $a'\beta$,

$$a'b(n) = a'\{X(n)\}^{+}y(n), \tag{4.2.8}$$

where $\{X(n)\}^{+}$ is the Moore-Penrose generalized inverse of $X(n)$, is weakly consistent if and only if

$$\lambda^{*}_{min}(X'(n)X(n)) \to \infty \quad \text{as} \quad n \to \infty.$$ [2]

In this section we determine necessary and sufficient conditions for the weak consistency of a class of linear estimators of β when $u(n)$ is $E(n,\Sigma(n))$; this class of estimators being those linear estimators which are unbiased whenever the first order moments of $u(n)$ exist.

Suppose $u(n)$ is an $E(n,\Sigma(n))$ random vector with joint density of the form[3]

$$g_n(u(n)) = |\Sigma(n)|^{-\frac{1}{2}}\phi_n(u'(n)\Sigma^{-1}(n)u(n)), \tag{4.2.9}$$

where

$$\phi_n : [o,\infty) \to [o,\infty)$$

and

$$\int_o^\infty r^{n-1}\phi_n(r^2)\,dr = \tfrac{1}{2}\Gamma(n/2)\pi^{-n/2}$$

for $n = 1, 2, \ldots$. For each value of n, $u(n)$ can be viewed as an n-dimensional sample from a spherically invariant stochastic

---

2. $\lambda^{*}_{min}(X'(n)X(n))$ denotes the smallest positive eigenvalue of $X'(n)X(n)$. When $X(n)$ is of full rank for large n, this condition is equivalent to (4.2.5).

3. Property VIII implies that a sufficient condition for $u(n)$ to have a joint density function is that $Pr(u(n+1)=0)=0$.

process and therefore, by property IX, ϕ_n has the form

$$\phi_n(\cdot) = \int_o^\infty (2\pi\tau^2)^{-n/2}\exp(-\cdot/2\tau^2)\,dF(\tau),$$

where $F(\cdot)$ is a distribution function with support $[o,\infty)$, $F(\cdot)$ being common for all values of n. A necessary and sufficient condition for the existence of the k^{th} order moment of $u(n)$ is that

$$\int_o^\infty\int_o^\infty r^{k+n-1}(2\pi\tau^2)^{-n/2}\exp(-r^2/2\tau^2)\,dF(\tau)\,dr < \infty$$

which can be shown to be equivalent to

$$\int_o^\infty \tau^k dF(\tau) < \infty$$

and hence independent of n. Therefore, if the first order moments of $u(n)$ exist for any value of n, they exist for all positive values of n.

When the first order moments of $u(n)$ do exist, any linear estimator of β of the form

(4.2.10) $$\hat{\beta}(n) = B(n)y(n),$$

where $B(n)$ is a $k \times n$ nonstochastic matrix, is unbiased if and only if[4]

(4.2.11) $$B(n)X(n) = I_k.$$

On the other hand, when the first order moments of $u(n)$ do not exist, all linear estimators have infinite means and consequently

---

4. This follows directly from the relationship $E(\hat{\beta}(n)) = B(n)X(n)\beta$.

are not unbiased in the classical sense. In order to allow for this situation, we study the class of linear estimators of the form (4.2.10) for which (4.2.11) holds, instead of confining our attention to linear estimators which are unbiased in the classical sense.

If (4.2.11) holds,

$$\hat{\beta}(n) - \beta = B(n)u(n) \tag{4.2.12}$$

and therefore by property III,

$$\delta(n) = \hat{\beta}(n) - \beta$$

is distributed $E(k,\Omega(n))$ with joint density function

$$h(\delta(n)) = |\Omega(n)|^{-\frac{1}{2}}\phi_k(\delta'(n)\Omega^{-1}(n)\delta(n)),$$

where

$$\Omega(n) = B(n)\Sigma(n)B'(n).$$

Theorem 4.1. Suppose the disturbance vector of the linear regression model (4.2.2) is an $E(n,\Sigma(n))$ random vector[5] for $n = k, k+1, \ldots$.

(i) If $\hat{\beta}(n)$ is a linear estimator of β of the form (4.2.10) such that (4.2.11) holds for all $n \geq k$, then it is a weakly consistent estimator of β if and only if

$$\lim_{n\to\infty} \Omega(n) = 0. \tag{4.2.13}$$

---

5. This theorem does not assume that $u(n)$, $n = k, k+1, \ldots$ have joint density functions. Note that if

$$\Pr(u(n)=0) = 0, \qquad n = k, k+1, \ldots,$$

property VIII implies that joint density functions do exist.

(ii) If both (4.2.6) and (4.2.7) hold then (4.2.13) holds if and only if

$$\lim_{n\to\infty} B(n)B'(n) = 0. \tag{4.2.14}$$

Proof: (i) Since $\{u_t;\ t=1,\ 2,\ \ldots\}$ is a spherically invariant stochastic process, property IX implies that the characteristic function of $u(n)$ can be expressed as:

$$\begin{aligned}\psi_{u(n)}(s) &= E\{\exp(is'u(n))\}\\ &= \int_o^\infty \exp(-\nu^2 s'\Sigma(n)s/2)\,dF(\nu),\end{aligned}$$

where s is an n-dimensional vector and $F(\cdot)$ is a probability distribution function with support $[o,\infty)$ and independent of n.

The characteristic function of $\delta(n)$, therefore, can be written as

$$\begin{aligned}\psi_{\delta(n)}(r) &= E[\exp\{ir'\delta(n)\}]\\ &= E[\exp\{ir'B(n)u(n)\}]\\ &= \psi_{u(n)}(B'(n)r)\\ &= \int_o^\infty \exp(-\nu^2 r'\Omega(n)r/2)\,dF(\nu),\end{aligned}$$

where r is a k-dimensional vector.

$\delta(n) \to 0$ in probability if and only if

$$\lim_{n\to\infty} F_{\delta(n)}(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x > 0, \end{cases}$$

where $F_{\delta(n)}(\cdot)$ is the probability distribution of $\delta(n)$. Therefore, by the Lévy-Cramér theorem,[6] $\delta(n) \to 0$ in probability if and only if for all k-dimensional vectors, r,

$$\psi_{\delta(n)}(r) \to 1 \quad \text{as} \quad n \to \infty.$$

Hence, all we need show is that

$$\lim_{n\to\infty} \psi_{\delta(n)}(r) = 1 \tag{4.2.15}$$

for all k-dimensional vectors, r, if and only if (4.2.13) holds.

If (4.2.13) holds then

$$\lim_{n\to\infty} \exp\{-\nu^2 r'\Omega(n)r/2\} = 1$$

for any k-dimensional vector r and any $\nu \in [o,\infty)$. Since for any k-dimensional vector r

$$0 < \exp\{-\nu^2 r'\Omega(n)r/2\} \leqslant 1, \tag{4.2.16}$$

Lebesgue's dominated convergence theorem implies that

$$\begin{aligned}\lim_{n\to\infty} \psi_{\delta(n)}(r) &= \lim_{n\to\infty} \int_o^\infty \exp\{-\nu^2 r'\Omega(n)r/2\}dF(\nu)\\ &= 1.\end{aligned}$$

Suppose (4.2.15) holds but $\Omega(n)$ does not converge to 0 as $n \to \infty$. Then there exists a subsequence $\{\tilde{n}\}$ such that for some r, either

$$r'\Omega(\tilde{n})r \to \omega > 0$$

---

6. See Fisz (1963, p.188).

or

$$r'\Omega(\tilde{n})r \to \infty$$

as $\tilde{n} \to \infty$. (4.2.16) and Lebesgue's dominated convergence theorem imply

$$\lim_{\tilde{n}\to\infty} \psi_{\delta(\tilde{n})}(r) < 1;$$

a contradiction. Therefore (4.2.15) implies (4.2.13).

(ii) For any value of $n > k$, note that (4.2.11) implies $\text{rank}(B(n)) = k$. Since

$$\lambda_{max}(\Sigma(n)) = \max_{x\neq o} \frac{x'\Sigma(n)x}{x'x}$$

and

$$\lambda_{min}(\Sigma(n)) = \min_{x\neq o} \frac{x'\Sigma(n)x}{x'x},$$

x being an n-dimensional vector,[7] then for all non-zero, k-dimensional vectors, r,

$$\begin{aligned}
&\lambda_{max}(\Sigma(n))r'B(n)B'(n)r \\
&\geq r'\Omega(n)r \\
&\geq \lambda_{min}(\Sigma(n))r'B(n)B'(n)r \\
&> 0.
\end{aligned}$$

Hence under (4.2.6) and (4.2.7), (4.2.13) holds if and only if (4.2.14) holds.

---

7. For a proof of this result see Graybill (1969, p.309).

Note that when (4.2.6) holds and

(4.2.17) $$\inf_{n\geq 1}\{\lambda_{\min}(\Sigma(n))\} = 0,$$

(4.2.14) is a sufficient condition for the weak consistency of $\hat{\beta}(n)$.

Theorem 4.1 is a very general result. The following corollaries demonstrate its application to some of the better known estimators of β in (4.2.2).

Corollary 4.1.1 (i) If $u(n)$ is $E(n,\Sigma(n))$ for $n = k, k+1, \ldots,$ the GLS estimator of β,

(4.2.18) $$\tilde{\beta}(n) = (X'(n)\Sigma^{-1}(n)X(n))^{-1}X'(n)\Sigma^{-1}(n)y(n)$$

is weakly consistent if and only if

(4.2.19) $$\lim_{n\to\infty}(X'(n)\Sigma^{-1}(n)X(n))^{-1} = 0.$$

(ii) If both (4.2.6) and (4.2.7) hold then (4.2.19) holds if and only if (4.2.5) holds.

Proof: (i) Follows immediately from Theorem 4.1 (i).

(ii) First note that (4.2.5) is equivalent[8] to

$$r'X'(n)X(n)r \to \infty \quad \text{as} \quad n \to \infty$$

for every k-dimensional vector $r \neq 0$. Similarly (4.2.19) is equivalent to

$$r'X'(n)\Sigma^{-1}(n)X(n)r \to \infty \quad \text{as} \quad n \to \infty$$

for every k-dimensional vector $r \neq 0$.

---

8. See Eicker (1963) and Anderson and Taylor (1976).

For any such r,

$$r'X'(n)X(n)r\lambda_{max}(\Sigma^{-1}(n)) \geqslant r'X'(n)\Sigma^{-1}(n)X(n)r$$
$$\geqslant r'X'(n)X(n)r\lambda_{min}(\Sigma^{-1}(n))$$

or equivalently since $\Sigma(n)$ is positive definite,

$$r'X'(n)X(n)r/\lambda_{min}(\Sigma(n)) \geqslant r'X'(n)\Sigma^{-1}(n)X(n)r$$
$$\geqslant r'X'(n)X(n)r/\lambda_{max}(\Sigma(n)).$$

Taking limits we get for any $r \neq 0$,

$$\lim_{n\to\infty}[r'X'(n)X(n)r]/\inf_{n\geqslant 1}\{\lambda_{min}(\Sigma(n))\}$$
$$\geqslant \lim_{n\to\infty}[r'X'(n)\Sigma^{-1}(n)X(n)r]$$
$$\geqslant \lim_{n\to\infty}[r'X'(n)X(n)r]/\sup_{n\geqslant 1}\{\lambda_{max}(\Sigma(n))\}$$

Hence $\lim_{n\to\infty}[r'X'(n)X(n)r] = \infty$ for all $r \neq 0$ if and only if $\lim_{n\to\infty}[r'X'(n)\Sigma^{-1}(n)X(n)r] = \infty$ for all $r \neq 0$ as required.

Corollary 4.1.2 (i) If $u(n)$ is $E(n,\Sigma(n))$ for $n = k, k + 1, \ldots,$ the GLS estimator with the disturbance covariance matrix (incorrectly) assumed to be $\Theta(n)$ is weakly consistent if and only if

(4.2.20) $$\lim_{n\to\infty}(X'(n)\Theta^{-1}(n)X(n))^{-1}X'(n)\Theta^{-1}(n)\Sigma(n)\Theta^{-1}(n)X(n)(X'(n)\Theta^{-1}(n)X(n))^{-1} = 0.$$

(ii) If

(4.2.21) $$\sup_{n\geqslant 1}\{\lambda_{max}(\Sigma(n)\Theta^{-1}(n))\} < \infty$$

and

$$\inf_{n\geq 1}\{\lambda_{min}(\Sigma(n)\Theta^{-1}(n))\} > 0, \tag{4.2.22}$$

then (4.2.20) holds if and only if

$$\lim_{n\to\infty}(X'(n)\Theta^{-1}(n)X(n))^{-1} = 0.$$

(iii) If in addition to (4.2.21) and (4.2.22),

$$\sup_{n\geq 1}\{\lambda_{max}(\Theta(n))\} < \infty$$

and

$$\inf_{n\geq 1}\{\lambda_{min}(\Theta(n))\} > 0,$$

then (4.2.20) holds if and only if (4.2.5) holds.

<u>Proof</u>: (i) Follows immediately from Theorem 4.1.

(ii) (4.2.20) can be written as

$$\lim_{n\to\infty} B(n)\Gamma(n)B'(n) = 0 \tag{4.2.23}$$

where

$$B(n) = (X'(n)\Theta^{-1}(n)X(n))^{-1}X'(n)\Theta^{-\frac{1}{2}}(n)$$

and

$$\Gamma(n) = \Theta^{-\frac{1}{2}}(n)\Sigma(n)\Theta^{-\frac{1}{2}}(n).$$

The required result can be obtained by applying the proof of Theorem 4.1 (ii) to (4.2.23) in place of (4.2.13) and noting that the characteristic roots of $\Theta^{-\frac{1}{2}}(n)\Sigma(n)\Theta^{-\frac{1}{2}}(n)$ and $\Sigma(n)\Theta^{-1}(n)$ correspond.

(iii) Follows immediately from Corollary 4.1.1 (ii).

Corollary 4.1.3 (i) If $u(n)$ is $E(n,\Sigma(n))$, $n = k, k+1, \ldots,$ the OLS estimator (4.2.4) is weakly consistent[9] if and only if

$$\text{(4.2.24)} \quad \lim_{n\to\infty}(X'(n)X(n))^{-1}X'(n)\Sigma(n)X(n)(X'(n)X(n))^{-1} = 0.$$

(ii) If both (4.2.6) and (4.2.7) hold then (4.2.24) holds if and only if (4.2.5) holds.

Theorem 4.1 can also be applied to Sargan's (1959) general instrumental variable estimator assuming the design matrix of (4.2.2) is non-stochastic. This assumption is not particularly realistic since the instrumental variable estimator is best used in situations where the design matrix is stochastic. The following corollary is presented merely to help demonstrate the versatility of Theorem 4.1.

Corollary 4.1.4. If $u(n)$ is $E(n,\Sigma(n))$, $n = k, k+1, \ldots$ and if both (4.2.6) and (4.2.7) hold then the general instrumental variable estimator

$$\tilde{b}(n) = (X'(n)N(n)X(n))^{-1}X'(n)N(n)y(n)$$

is weakly consistent if and only if

$$\lim_{n\to\infty}(X'(n)N(n)X(n))^{-1} = 0$$

---

9. A special case of this result has been proved by Jensen (1978). In a note that appeared after this chapter was first drafted, he demonstrated that in the model $y(n) = \theta\ell(n) + u(n)$ where $\ell(n) = (1,\ldots,1)'$ and $u(n)$ takes a Cauchy spherically symmetric distribution, the OLS estimator of the scalar θ, namely $\frac{1}{n}\sum_{i=1}^{n} y_i(n)$, is weakly consistent.

where $N(n) = Z(n)(Z'(n)Z(n))^{-1}Z'(n)$ and $Z(n)$ is an $(n \times g)$ matrix of observations on g instruments with $g \geqslant k$.

If under (4.2.2), the set of j linear restrictions,

$$R\beta = r, \tag{4.2.25}$$

holds, where R is a known $j \times k$ matrix of rank $j < k$ and r is a known j-dimensional vector, the restricted least squares estimator,

$$\hat{b}(n) = b(n) + (X'(n)X(n))^{-1}R'[R(X'(n)X(n))^{-1}R']^{-1}(r-Rb(n)), \tag{4.2.26}$$

where $b(n)$ is the OLS estimator of β, (4.2.4), is not a linear estimator of the form of (4.2.10) and hence Theorem 4.1 cannot be applied. However,

$$\hat{b}(n) - \beta = A(n)(X'(n)X(n))^{-1}X'(n)u(n),$$

where

$$A(n) = I_k - (X'(n)X(n))^{-1}R'[R(X'(n)X(n))^{-1}R']^{-1}R.$$

Hence if $u(n)$ is an $E(n,\Sigma(n))$ random vector for $n = k, k+1, \ldots$, then by property III

$$\hat{\delta}(n) = \hat{b}(n) - \beta$$

is distributed $E(k,\hat{\Omega}(n))$, where

$$\hat{\Omega}(n) = A(n)(X'(n)X(n))^{-1}X'(n)\Sigma(n)X(n)(X'(n)X(n))^{-1}A'(n).$$

By applying the arguments contained in the proof of Theorem 4.1 we have:

Corollary 4.1.5. If under (4.2.2), $u(n)$ is $E(n,\Sigma(n))$, $n = k, k+1, \ldots$, such that (4.2.6) and (4.2.7) hold and if the set of j linear restrictions (4.2.25) hold then the restricted least squares estimator (4.2.26) is weakly consistent if and only if

$$\lim_{n\to\infty} A(n)(X'(n)X(n))^{-1} = 0.$$

An interesting feature of Theorem 4.1 is that it shows a whole class of estimators to be weakly consistent without requiring the existence of first and/or second order moments of the joint distribution of the disturbances. For example, the OLS estimator of β is weakly consistent when the disturbances of (4.2.2) take a spherically symmetric Cauchy distribution. At first sight this may appear to contradict Poisson's (1824) result that for the Cauchy distribution the mean of a random sample follows the same distribution law as the parent distribution. Note that property IV implies that if u is Cauchy $E(n,I_n)$, its components are dependent. It is this dependency which has allowed us to establish weak consistency of the OLS estimator.

Three alternative forms of condition (4.2.13) are

(i) $\lambda_{max}(\Omega(n)) \to 0$ as $n \to \infty$,

(ii) $\lambda_{min}(\Omega^{-1}(n)) \to \infty$ as $n \to \infty$ and

(iii) $\gamma'\Omega(n)\gamma \to 0$ for every k-dimensional vector $\gamma \neq 0$ as $n \to \infty$.

Similar alternative forms apply for condition (4.2.14).

Finally consider the case where the components of $u(n)$ are generated by the stationary first-order autoregressive scheme

(4.2.27) $$u_t = \rho u_{t-1} + e_t \qquad |\rho|<1,\ t=1,2,\ldots,$$

where

$$e(n) = (e_1,\ldots,e_n)'$$

is distributed $E(n,I_n)$. By property III u(n), is $E(n,\Sigma(n))$ with $\Sigma(n)$ being of the form

(4.2.28) $$\Sigma(n) = [(1-\rho)^2 I_n + \rho A_1 + \rho(1-\rho)C_1]^{-1},$$

where A_1 is the n x n matrix

(4.2.29) $$A_1 = \begin{bmatrix} 1 & -1 & 0 \ldots\ldots & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & 0 & 0 \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 0 & 0 & 0 & 2 & -1 \\ 0 & 0 & 0 \ldots\ldots & -1 & 1 \end{bmatrix}$$

and C_1 is the n x n matrix

(4.2.30) $$C_1 = \begin{bmatrix} 1 & 0 \ldots\ldots & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & 0 & 0 & 0 \\ 0 & 0 \ldots\ldots & 0 & 1 \end{bmatrix}.$$

$\Sigma(n)$ can be approximated by

(4.2.31) $$\bar{\Sigma}(n) = [(1-\rho)^2 I_n + \rho A_1]^{-1}.$$

Note that the degree of approximation decreases as n increases. The characteristic roots of $\bar{\Sigma}(n)$ are

$$[(1-\rho)^2 + 2\rho\{1-\cos(\pi(i-1)/n)\}]^{-1} \qquad i=1,2,\ldots,n,$$

which are bounded by $(1-\rho)^{-2}$ and $(1+\rho)^{-2}$, so that for any given $\rho \in (-1,1)$, (4.2.6) and (4.2.7) clearly hold for $\bar{\Sigma}(n)$. We conjecture that (4.2.6) and (4.2.7) also hold for $\Sigma(n)$ when $\rho \in (-1,1)$.

If our conjecture is true, (4.2.5) would be a necessary and sufficient condition for the OLS estimator, b(n), to be weakly consistent when the regression disturbances follow the first-order autoregressive scheme (4.2.27) in which the errors, e_t, are jointly spherically symmetric. Somewhat surprisingly, (4.2.5) would also be a necessary and sufficient condition for the GLS estimator of β to be weakly consistent either when the value taken by ρ is known and used to obtain $\Sigma(n)$ or when any weakly consistent estimator of ρ is used in (4.2.28) to estimate $\Sigma(n)$.

3. STRONG CONSISTENCY OF LEAST SQUARES REGRESSION ESTIMATORS

Strong consistency of any sequence of estimators is a more stringent property than weak consistency and is a much harder property to establish. It is only comparatively recently that sufficient conditions have been found for strong consistency of least squares regression estimators.

Jennrich (1969) considered the non-linear regression model

$$y_t = f_t(\theta_o) + u_t, \qquad t=1,2,\ldots .$$

Assuming the disturbance terms to be identically, independently distributed with zero mean and finite variance, he found conditions on the sequence of functions $\{f_t\}$ such that the least squares estimator of Θ_o is strongly consistent. For the linear regression model, (4.2.2), these conditions are extremely restrictive. For example, when $k = 1$, they are

(i) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i^2$ exists and

(ii) there exists an integer m such that for all $n > m$,
$\frac{1}{n} \sum_{i=1}^{n} x_i^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i^2$.

When the disturbances of (4.2.1), u_t, $t = 1, 2, \ldots,$ form a sequence of independent random variables with mean zero and

$$\sup_{t\geq 1} E(u_t^2) < \infty,$$

Drygas (1976) showed that a sufficient condition[10] for strong consistency of the OLS estimator, $b(n)$, is that there exists a sequence of positive constants, $\tau_n \to \infty$, and a finite positive definite matrix, Θ, such that

$$(X'(n)X(n))/\tau_n \to \Theta$$

as $n \to \infty$. For $k = 1$ and when $x_1, x_2, \ldots$ are random variables such that

$$E(u_t|x_1,\ldots) = 0, \qquad t=1,2,\ldots,$$

$E(u_t^2|x_1,\ldots)$, $t = 1, 2, \ldots$ are bounded almost surely and $u_1, u_2, \ldots$ given $x_1, x_2, \ldots$ are independent, he also found that

---

10. Drygas presented these conditions as being sufficient for weaker more complex conditions to hold. Under this latter set of conditions, he was able to establish strong consistency of the least squares estimator, (4.2.8), of the estimable linear combination, $a'\beta$, when $X(n)$ is not necessarily of full rank.

$$\sum_{i=1}^{\infty} x_i^2 = \infty$$

with probability one is a necessary and sufficient condition for the strong consistency of $b(n)$.

Anderson and Taylor (1976) demonstrated that when $u(n)$ is distributed $N(0,\sigma^2 I_n)$, $b(n)$ is a strongly consistent estimator of β if and only if (4.2.5) holds. More recently, Lai and Robbins (1977) considered the linear regression model,

$$y_t = \beta_1 + \beta_2 x_{2t} + u_t, \qquad t=1,2,\ldots,$$

where u_1, u_2, ... are independent identically distributed random variables such that

$$E(u_t) = 0,\ E(u_t^2) = \sigma^2, \qquad 0<\sigma^2<\infty,$$

and $E[u_t^2 \log(1+|u_t|)^r] < \infty$ for some $r > 1$, $t = 1, 2, \ldots$. They obtained

$$\lim_{n\to\infty} \sum_{t=1}^{n} (x_{2t}-\tilde{x}_{2n})^2 = \infty,$$

where $\tilde{x}_{2n} = \frac{1}{n} \sum_{t=1}^{n} x_{2t}$, as a sufficient condition for strong consistency of $b(n)$.

In this section, we attempt first to fill a gap in this literature by finding sufficient conditions for strong consistency of $b(n)$, when the only assumption made about the regression disturbances is that (4.2.3) holds, i.e. the disturbances are uncorrelated with bounded variances. Then the assumption that the disturbances are elliptically symmetric is added, allowing (4.2.5) to be established as a sufficient condition for strong consistency of $b(n)$ and a necessary condition if

(4.3.1) $$\inf_{t\geq 1} \sigma_t^2 > 0.$$

In both cases analogous results are obtained for the GLS estimator.

Theorem 4.2. With respect to the linear regression model (4.2.2), let $A(n) = X'(n)X(n)$. If (4.2.3) holds and the sequence

(4.3.2) $$\{(\log^{2+\varepsilon} n)A^{-1}(n),\ n=k,k+1,\ldots\}$$

converges for some $\varepsilon > 0$, as $n \to \infty$, then $b(n)$, given by (4.2.4), is a strongly consistent estimator of β.

In proving Theorem 4.2 we shall make use of the following results:

Lemma 4.1. If $\sum_{i=1}^{n} d_i$ is an arbitrary divergent series of positive terms, then

$$\sum_{i=1}^{n} d_i/(\sum_{j=1}^{i} d_j)^{\alpha}$$

is convergent for $\alpha > 1$ and divergent for $\alpha \leq 1$.
[For a proof see Knopp (1956, p.125).]

Lemma 4.2. (The Rademacher-Menchoff fundamental convergence theorem for orthogonal random variables.) If $\{z_i, i=1,2,\ldots\}$ is a sequence of random variables such that

$$E[z_i z_j] = 0 \quad \text{for all} \quad i \neq j$$

and

$$\sum_{n=1}^{\infty} (\log^2 n)E[z_n^2] < \infty,$$

then $\sum_{i=1}^{n} z_i$ converges with probability one.
[For a proof see Stout (1974, p.20).]

Lemma 4.3. (Kronecker's Lemma.) Let $\{z_i, i=1,2,\ldots\}$ be an arbitrary numerical sequence and $\{\rho_i, i=1,2,\ldots\}$ a numerical sequence such that

$$0 < \rho_1 \leqslant \rho_2 \leqslant \ldots \to \infty .$$

If $\sum_{i=1}^{\infty} \rho_i^{-1} z_i < \infty$ then

$$\rho_n^{-1} \sum_{i=1}^{n} z_i \to 0 \quad \text{as} \quad n \to \infty .$$

[For a proof see Feller (1971, p.239).]

Proof of Theorem 4.2.

We shall show that if (4.3.2) holds, $b(n) - \beta \to 0$ with probability one as $n \to \infty$, making use of the relationship

$$b(n) - \beta = A^{-1}(n)X'(n)u(n). \tag{4.3.3}$$

First consider the case of $k = 1$. Then x_t is a scalar. Let

$$\lambda_n = \sum_{t=1}^{n} x_t^2 ,$$

so that (4.3.2) becomes,

$$\{(\log^{2+\varepsilon} n)\lambda_n^{-1},\ n=1,2,\ldots\} \tag{4.3.4}$$

converges for some $\varepsilon > 0$ as $n \to \infty$, while (4.3.3) can be written as

$$b(n) - \beta = \sum_{t=1}^{n} x_t u_t / \lambda_n .$$

Since the function $\eta \to \eta^{\nu}$ is continuous [11] for any real ν,

---

11. The function $f:(-\infty,\infty) \to (-\infty,\infty)$ is continuous if and only if for every sequence $\{z_n, n=1,2,\ldots\}$ converging to $z \in (-\infty,\infty)$, the sequence $\{f(z_n)\}$ converges to $f(z)$. See White (1968, p.62).

(4.3.4) implies that the sequence

$$\{(\log^2 n)\lambda_n^{-\alpha},\ n=1,2,\ldots\}$$

converges for $\alpha = 2/(2+\varepsilon)$, i.e. $0 < \alpha < 1$, as $n \to \infty$. (4.3.4) also implies

$$\lambda_n^{-1} \to 0 \quad \text{as} \quad n \to \infty,$$

hence by Lemma 4.1,

$$\sum_{n=1}^{\infty} (x_n^2/\lambda_n^{2-\alpha}) < \infty.$$

That

$$(4.3.5) \qquad \sum_{n=1}^{\infty} (\log^2 n)\, x_n^2/\lambda_n^2 < \infty$$

follows upon application of the comparison test.

Let

$$z_n = x_n u_n/\lambda_n.$$

Then $\{z_n, n=1,2,\ldots\}$ is a sequence of uncorrelated random variables with mean zero and (4.3.5) implies

$$\sum_{n=1}^{\infty} (\log^2 n) E(z_n^2) < \infty.$$

Therefore,

$$\sum_{n=1}^{\infty} z_n = \sum_{n=1}^{\infty} x_n u_n/\lambda_n < \infty$$

with probability one by Lemma 4.2 while Lemma 4.3 implies

$$\sum_{t=1}^{n} x_t u_t/\lambda_n = b(n) - \beta \to 0$$

with probability one as $n \to \infty$.

For $k > 1$, without loss of generality, we consider just the first component of $b(n) - \beta$. Let

$$\beta = \begin{bmatrix} \beta_1 \\ \bar{\beta} \end{bmatrix}, \quad x_t = \begin{bmatrix} x_{1t} \\ \bar{x}_t \end{bmatrix}, \quad b(n) = \begin{bmatrix} b_1(n) \\ \bar{b}(n) \end{bmatrix},$$

$$X(n) = [x_{(1)}(n) \vdots \bar{X}(n)], \quad A(n) = \begin{bmatrix} a_{11}(n) & A_{12}(n) \\ A_{21}(n) & A_{22}(n) \end{bmatrix},$$

where β_1, x_{1t}, $b_1(n)$ and $a_{11}(n)$ are scalars, $\bar{\beta}$, $\bar{x}_t$, $A_{21}(n) = A_{12}'(n)$ are $(k-1)$-dimensional vectors, $x_{(1)}(n)$ is an n-dimensional vector whose components are x_{1t}, $t = 1, 2, \ldots, n$, $\bar{X}(n)$ is an $n \times (k-1)$ matrix with rows $\bar{x}_t'$, $t = 1, 2, \ldots, n$ and $A_{22}(n)$ is a $(k-1) \times (k-1)$ matrix. Also let

(4.3.6) $$\sigma^2 = \sup_{t \geq 1} \sigma_t^2 < \infty.$$

Since

$$\begin{aligned} &\sum_{t=1}^{n} (x_{1t} - A_{12}(n)A_{22}^{-1}(n)\bar{x}_t)^2 \\ &= (x_{(1)}'(n) - A_{12}(n)A_{22}^{-1}(n)\bar{X}'(n))(x_{(1)}'(n) - A_{12}(n)A_{22}^{-1}(n)\bar{X}'(n))' \\ &= a_{11}(n) - A_{12}(n)A_{22}^{-1}(n)A_{21}(n), \end{aligned}$$

$$b_1(n) - \beta_1 = \frac{(x_{(1)}'(n) - A_{12}(n)A_{22}^{-1}(n)\bar{X}'(n))u(n)}{(a_{11}(n) - A_{12}(n)A_{22}^{-1}(n)A_{21}(n))}$$

$$= \xi_n/\mu_n,$$

where $$\xi_n = \sum_{t=1}^{n} (x_{1t} - A_{12}(n)A_{22}^{-1}(n)\bar{x}_t)u_t,$$

$$\mu_n = \sum_{t=1}^{n} (x_{1t} - A_{12}(n)A_{22}^{-1}(n)\bar{x}_t)^2,$$

$$E(\xi_n) = 0 \quad \text{and}$$

$$
\begin{aligned}
&E[(\xi_{n+1} - \xi_n)^2] \\
\leqslant\ & \sum_{t=1}^{n} (A_{12}(n)A_{22}^{-1}(n)\bar{x}_t - A_{12}(n+1)A_{22}^{-1}(n+1)\bar{x}_t)^2\sigma^2 \\
&\qquad + (x_1(n+1) - A_{12}(n+1)A_{22}^{-1}(n+1)\bar{x}_{n+1})^2\sigma^2 \\
=\ & \{(A_{12}(n)A_{22}^{-1}(n)\bar{X}'(n) - A_{12}(n+1)A_{22}^{-1}(n+1)\bar{X}'(n))\cdot \\
&\quad (A_{12}(n)A_{22}^{-1}(n)\bar{X}'(n) - A_{12}(n+1)A_{22}^{-1}(n+1)\bar{X}'(n))' \\
&\qquad + (x_1(n+1) - A_{12}(n+1)A_{22}^{-1}(n+1)\bar{x}_{n+1})^2\}\sigma^2 \\
=\ & \{A_{12}(n)A_{22}^{-1}(n)A_{21}(n) - A_{12}(n+1)A_{22}^{-1}(n+1)A_{21}(n+1) + x_1^2(n+1)\}\sigma^2 \\
=\ & (\mu_{n+1} - \mu_n)\sigma^2 .
\end{aligned}
$$

Define

$$
\begin{aligned}
\gamma_1^2 &= \mu_k, \\
\gamma_t^2 &= \mu_{t+k-1} - \mu_{t+k-2}, \qquad && t=2,3,\ldots, \\
v_1 &= \xi_k/\gamma_1, \\
v_t &= (\xi_{t+k-1} - \xi_{t+k-2})/\gamma_t, \qquad && t=2,3,\ldots .
\end{aligned}
$$

The increment $\xi_{n+1} - \xi_n$ is uncorrelated[12] with $\xi_k, \xi_{k+1}, \ldots, \xi_n$ for $n = k, k+1, \ldots,$ hence $\{v_t, t=1,2,\ldots\}$ is a sequence of uncorrelated random variables with mean zero and variance

$$E(v_t^2) \leqslant \sigma^2.$$

Since

$$\sum_{t=1}^{n-k+1} \gamma_t^2 = \mu_n = a_{11}(n) - A_{12}(n)A_{22}^{-1}(n)A_{21}(n)$$

---

12. See Anderson and Taylor (1976, p.789).

is the reciprocal of the upper left hand corner of $A^{-1}(n)$, the sequence

$$\{\log^{2+\varepsilon} n(\sum_{t=1}^{n-k+1} \gamma_t^2)^{-1}, \ n=1,2,\ldots\}$$

is convergent as $n \to \infty$ for some $\varepsilon > 0$. The argument used for $k = 1$ can be applied to show that

$$b_1(n) - \beta_1 = \sum_{t=1}^{n-k+1} \gamma_t v_t \Big/ \sum_{t=1}^{n-k+1} \gamma_t^2$$

converges to 0 with probability 1, as $n \to \infty$.

It is interesting to note that heteroscedasticity does not prevent the OLS estimator from being strongly consistent, provided the error variances are bounded.

Corollary 4.2.1. If in the linear regression model (4.2.2) the disturbance vector $u(n)$ has zero mean and covariance matrix $\sigma^2\Sigma(n)$, where $\Sigma(n)$ is a positive definite, $n \times n$ matrix for $n = 1, 2, \ldots$, then if the sequence

$$\{\log^{2+\varepsilon} n[X'(n)\Sigma^{-1}(n)X(n)]^{-1}, \ n=k,k+1,\ldots\}$$

converges for some $\varepsilon > 0$ as $n \to \infty$, the GLS estimator, $\tilde{\beta}(n)$, given by (4.2.18), is a strongly consistent estimator of β.

Proof: For each sequence of covariance matrices

$$\{\Sigma(n), \ n=1,2,\ldots\},$$

there exists a sequence of triangular, non-singular, matrices

$$\{C(n), \ n=1,2,\ldots\}, \tag{4.3.7}$$

such that for each n,

$$C(n)\Sigma(n)C'(n) = I_n \tag{4.3.8}$$

and $C(n)$ is of the form

$$C(n) = \begin{bmatrix} C(m) & 0 \\ C_{21}^m(n) & C_{22}^m(n) \end{bmatrix} \quad \text{for all } m < n, \tag{4.3.9}$$

where $C_{21}^m(n)$ is $(n-m) \times m$, $C_{22}^m(n)$ is $(n-m) \times (n-m)$ and 0 is an $m \times (n-m)$ matrix of zeros.

In order to confirm this, suppose there exist non-singular matrices $C(n)$ of the form (4.3.9) such that (4.3.8) holds for $n = 1, \ldots, j-1$. Partition $\Sigma(j)$ and $C(j)$ as

$$\Sigma(j) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \sigma_{jj}^2 \end{bmatrix} \quad \text{and} \quad C(j) = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & c_{22} \end{bmatrix},$$

where Σ_{11} and C_{11} are $(j-1) \times (j-1)$, C_{21}', C_{12} and $\Sigma_{12} = \Sigma_{21}'$ are $(j-1) \times 1$ while σ_{jj}^2 and c_{22} are scalars. For (4.3.8) to hold for $C(j)$ implies

$$\begin{aligned}
C_{11}\Sigma_{11}C_{11}' + C_{12}\Sigma_{21}C_{11}' + C_{11}\Sigma_{12}C_{12}' + \sigma_{jj}^2 C_{12}C_{12}' &= I_{(j-1)} \\
C_{21}\Sigma_{11}C_{11}' + c_{22}\Sigma_{21}C_{11}' + C_{21}\Sigma_{12}C_{12}' + \sigma_{jj}^2 c_{22}C_{12}' &= 0 \\
C_{21}\Sigma_{11}C_{21}' + c_{22}\Sigma_{21}C_{21}' + c_{22}C_{21}\Sigma_{12} + \sigma_{jj}^2 c_{22}^2 \quad &= 1.
\end{aligned}$$

It is easily verified that a solution to this set of equations is

$$\begin{aligned}
C_{11} &= C(j-1) \\
C_{12} &= 0 \\
C_{21} &= -\Sigma_{21}\Sigma_{11}^{-1}/\zeta \\
c_{22} &= {}^1/_\zeta,
\end{aligned}$$

where

$$1/\zeta = (\sigma_{jj}^2 - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-\frac{1}{2}},$$

i.e. the square root of the $(n,n)^{th}$ component of $\Sigma^{-1}(j)$. Since $\Sigma^{-1}(j)$ is positive definite, $\zeta > 0$, and $C(j)$ is non-singular. Therefore the existence of the sequence of non-singular matrices, (4.3.7), is guaranteed by the fact that such a matrix $C(n)$ exists for $n = 1$.

After transforming (4.2.2) by premultiplying by $C(n)$ for $n = 1, 2, \ldots$, Theorem 4.2 can be applied since (4.3.9) assures that $C(n)y(n)$ for $n = m, m+1, \ldots$ have the same first m components and that the first m rows of $C(n)X(n)$ for $n = m, m+1, \ldots$ are identical. Note that (4.3.8) and the non-singularity of $C(n)$ imply

$$C'(n)C(n) = \Sigma^{-1}(n).$$

We now turn our attention to the strong consistency of the OLS estimator of β in (4.2.2) when the regression disturbances are elliptically symmetric.

Theorem 4.3. With respect to the linear regression model (4.2.2), let $A(n) = X'(n)X(n)$. If the disturbance vector, $u(n)$, is distributed $E(n,\Sigma(n))$ for $n = 1, 2, \ldots$ such that (4.2.3) holds then

$$A^{-1}(n) \to 0 \quad \text{as} \quad n \to \infty \tag{4.3.10}$$

is a sufficient condition for the strong consistency of the OLS estimator of β, $b(n)$. If, moreover, (4.3.1) holds, then (4.3.10) is also a necessary condition for strong consistency.[13]

In proving Theorem 4.3 we shall use the following Lemma.

Lemma 4.4. Let $\{z_n, n=1,2,...\}$ be a sequence of random variables such that

$$E[z_n | z_1,...,z_{n-1}] = 0$$

for all n. If $\{\rho_i, i=1,2,...\}$ is a sequence of scalars such that

$$0 < \rho_1 \leqslant \rho_2 \leqslant ... \rightarrow \infty \tag{4.3.11}$$

and

$$\sum_{n=1}^{\infty} \rho_n^{-2} E(z_n^2) < \infty, \tag{4.3.12}$$

then

$$\rho_n^{-1} \sum_{i=1}^{n} z_i \rightarrow 0$$

with probability one as $n \rightarrow \infty$.

[For a proof see Feller (1971, p.243).]

Proof of Theorem 4.3

Sufficiency: We shall show that if (4.3.10) holds, $b(n) - \beta \rightarrow 0$ with probability one as $n \rightarrow \infty$.

---

13. Since this chapter was first drafted, an article by Lai, Robbins and Wei (1978) has been brought to my attention. It contains a proof of the result that (4.3.10) is a sufficient condition for the strong consistency of $b(n)$, when the disturbances of (4.2.1), $\{u_t, t=1,2,...\}$, form a martingale difference sequence and are such that (4.2.3) holds. Since under the conditions of Theorem 4.3, $\{u_t, t=1,2,...\}$ is a martingale difference sequence, Lai et al. provide an alternative proof of the first part of this theorem.

First consider the case of $k = 1$. Then (4.3.10) becomes

$$(\sum_{t=1}^{n} x_t^2)^{-1} \to 0 \quad \text{as} \quad n \to \infty. \tag{4.3.13}$$

Let $z_n = x_n u_n$ and $\rho_n = \sum_{t=1}^{n} x_t^2$, $n=1,2,\ldots$. Property V implies

$$E(u_n | u_1,\ldots,u_{n-1}) = 0, \qquad n=1,2,\ldots,$$

and hence

$$E(z_n | z_1,\ldots,z_{n-1}) = 0, \qquad n=1,2,\ldots .$$

Further, (4.3.13) implies (4.3.11), hence

$$\sum_{n=1}^{\infty} \rho_n^{-2} E(z_n^2) \leqslant \sigma^2 \sum_{n=1}^{\infty} \{ x_n^2 / (\sum_{t=1}^{n} x_t^2)^2 \} < \infty$$

by Lemma 4.1, where σ^2 is given by (4.3.6). Application of Lemma 4.4 yields

$$b(n) - \beta = \sum_{t=1}^{n} x_t u_t \Big/ \sum_{t=1}^{n} x_t^2 \to 0$$

with probability one as $n \to \infty$.

For $k > 1$, without loss of generality we shall consider just the first component of $b(n) - \beta$. As in the latter half of the proof of Theorem 4.2, let

$$\beta = \begin{bmatrix} \beta_1 \\ \bar{\beta} \end{bmatrix}, \qquad b(n) = \begin{bmatrix} b_1(n) \\ \bar{b}(n) \end{bmatrix} .$$

In that proof it was shown that $b_1(n) - \beta_1$ can be expressed as

$$b_1(n) - \beta_1 = \sum_{t=1}^{n-k+1} \gamma_t v_t \Big/ \sum_{t=1}^{n-k+1} \gamma_t^2,$$

where v_t is a linear transformation of $u(t)$ such that $\{v_t\}$

is a sequence of uncorrelated random variables with mean zero and variance $E(v_t^2) \leq \sigma^2$, and where $\{\gamma_t\}$ is a numerical sequence such that

$$(\sum_{t=1}^{n-k+1} \gamma_t^2)^{-1} \to 0 \quad \text{as} \quad n \to \infty.$$

Properties III and IV, together with (4.2.3) holding for $u(n)$, imply that $v(n) = (v_1, \ldots, v_n)'$ is an $E(n, \Theta(n))$ random vector such that (4.2.3) holds. Therefore the argument used for $k = 1$ can be applied to show that

$$b_1(n) - \beta_1 \to 0$$

with probability one as $n \to \infty$.

Necessity: Follows from part (ii) of Corollary 4.1.3 and the fact that strong consistency implies weak consistency.

Corollary 4.3.1. If the disturbance vector, $u(n)$, in the linear regression model (4.2.2) is distributed $E(n, \Sigma(n))$ with finite second order moments for $n = 1, 2, \ldots,$ then

$$[X'(n)\Sigma^{-1}(n)X(n)]^{-1} \to 0 \quad \text{as} \quad n \to \infty$$

is a necessary and sufficient condition for the strong consistency of the GLS estimator of β, $\tilde{\beta}(n)$, given by (4.2.18).

Proof: Follows from Theorem 4.3 and property III upon premultiplication of (4.2.2) by $C(n)$, where $\{C(n), n=1,2,\ldots\}$ is a sequence of triangular, nonsingular matrices of the form (4.3.9) such that for each n, (4.3.8) holds.

Strong consistency of $b(n)$ can be established without assuming the existence of second order moments of $u(n)$ as the following example shows. However, it does appear that the weakening of this assumption means that more stringent conditions on $X(n)$ are required for strong consistency to still hold.

For example, consider the case of (4.2.2) with $k = 1$. Condition (4.3.12) of Lemma 4.4 can be weakened to[14]

$$\sum_{n=1}^{\infty} \rho_n^{-1} E|z_n| < \infty .$$

Suppose $u(n)$ is $E(n,\Sigma(n))$, where $\Sigma(n)$ is a diagonal matrix for $n = 1, 2, \ldots$ and

$$E(u_t) = 0, \qquad t=1,2,\ldots,$$

$$\sup_{t\geq 1} E(|u_t|) < \infty .$$

It follows by a similar argument to that used in the proof of Theorem 4.3 that if in addition to (4.3.13),

$$\sum_{n=1}^{\infty} \{|x_n| \Big/ \sum_{t=1}^{n} x_t^2\} < \infty,$$

then $b(n)$ is a strongly consistent estimator of β.

4. ASYMPTOTIC DISTRIBUTIONS OF LEAST SQUARES REGRESSION ESTIMATORS

A typical approach to the subject of the asymptotic distribution of the OLS estimator in econometric textbooks,[15] is to prove that when

---

14. See Stout (1974, p.47).

15. For example, see Malinvaud (1970), Theil (1971) and Schmidt (1976).

the disturbances of (4.2.1), u_t, $t = 1, 2, \ldots$, are independently, identically distributed with zero mean and finite variance, σ^2, and when

$$(4.4.1) \qquad Q = \lim_{n\to\infty}\{X'(n)X(n)/n\}$$

is finite and non-singular, then the asymptotic distribution of $\sqrt{n}(b(n)-\beta)$ is $N(0,\sigma^2 Q^{-1})$, where $b(n)$ is the OLS estimator, (4.2.4). Eicker (1963) found necessary and sufficient conditions for the asymptotic normality of $\sqrt{n}(b(n)-\beta)$ in the case when the disturbances are independently but not necessarily identically distributed with mean zero and finite variances.

Asymptotic normality is one property of the OLS estimator which does not hold in general for spherically symmetric disturbances, the only exception being when $u(n)$ is multivariate normal. If $u(n)$ is $E(n,I_n)$ with a joint density function of the form (4.2.9) with $\Sigma(n) = I_n$, property III implies

$$\eta(n) = \sqrt{n}(b(n)-\beta)$$

is $E(k,\{X'(n)X(n)/n\}^{-1})$ with a joint density function of the form

$$h(r) = |X'(n)X(n)/n|^{\frac{1}{2}}\phi_k(r'X'(n)X(n)r/n),$$

where r is a k-dimensional vector. Therefore, when (4.4.1) holds, $\sqrt{n}(b(n)-\beta)$ has an $E(k,Q^{-1})$ asymptotic distribution with joint density function

$$h(r) = |Q|^{\frac{1}{2}}\phi_k(r'Qr),$$

which is a normal density if and only if (4.2.9) is a normal density.

At first sight this may appear to be in conflict with Eicker's result, but note that Eicker requires the disturbances to be independent while property IV implies that $N(0,\sigma^2 I_n)$ is the only $E(n,I_n)$ distribution with independent components. Clearly this dependency amongst the disturbances is preventing $\sqrt{n}b(n)$ from having an asymptotic normal distribution.

Similarly, when $u(n)$ is $E(n,\Sigma(n))$ with joint density (4.2.9) and

$$R = \lim_{n\to\infty} X'(n)\Sigma^{-1}(n)X(n)/n$$

is finite and non-singular,

$$\mu(n) = \sqrt{n}(\tilde{\beta}(n)-\beta),$$

where $\tilde{\beta}(n)$ is the GLS estimator (4.2.18), has an $E(k,R^{-1})$ asymptotic distribution with joint density

$$h(r) = |R|^{\frac{1}{2}}\phi_k(r'Rr).$$

5. AN ALTERNATIVE TO THE GAUSS-MARKOV THEOREM FOR ELLIPTICALLY SYMMETRIC DISTURBANCES

For the remainder of this chapter we shall write the linear regression model (4.2.2) as

$$y = X\beta + u. \tag{4.5.1}$$

We shall continue to assume that X is an $n \times k$, full rank, non-stochastic matrix.

If

$$(4.5.2) \qquad \begin{cases} E(u) = 0 \quad \text{and} \\ E(uu') = \sigma^2\Sigma, \end{cases}$$

where Σ is an $n \times n$ positive definite matrix, then a justification for the use of the GLS estimator,

$$(4.5.3) \qquad \tilde{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y,$$

as an estimator of β in (4.5.1) is provided by the Gauss-Markov theorem. This theorem states that $\tilde{\beta}$ is best within the class of linear unbiased estimators in the sense that every other linear unbiased estimator has a covariance matrix equal to the sum of the covariance matrix of $\tilde{\beta}$ and a positive semi-definite matrix.

The strength of this theorem depends upon whether or not one believes that linearity and unbiasedness are desirable properties that all estimators of β should have and also upon whether (4.5.2) holds. We do not intend to add to the already considerable literature on the former point. It remains a valid point of contention in Theorem 4.4 which is presented below.

When the disturbance vector of (4.5.1) is distributed $\mathcal{E}(n,\Sigma)$, the Gauss-Markov theorem is applicable only when the second order moments of u exist. As this cannot always be guaranteed, we are left with the question of whether an optimality result independent of the existence of second order moments of u holds for $\tilde{\beta}$ when the disturbances take an $\mathcal{E}(n,\Sigma)$ distribution. This section answers this question in the affirmative. We show that in the case of elliptically symmetric disturbances, the Gauss-Markov theorem can be replaced by a stronger optimality result which does not assume the existence of second order moments.

For the reasons expressed in Section 2, we shall confine our attention to the class of linear estimators of β of the form

$$\hat{\beta} = By \tag{4.5.4}$$

where B is a $k \times n$ nonstochastic matrix such that

$$BX = I_k . \tag{4.5.5}$$

It is pointless to compare covariance matrices of such estimators because

$$\hat{\beta} - \beta = Bu \tag{4.5.6}$$

and property III imply that if the second order moments of u do not exist those of $\hat{\beta}$ also will not exist. We need some other optimality criterion.

Intuitively, we would like estimates from our "optimum" estimator to be "close" to β with higher probability than those from any other member of the class of estimators under consideration. Translating this idea into mathematical terms, the optimum estimator would maximize

$$\Pr\{d(\hat{\beta}-\beta) < \varepsilon\}$$

for all real $\varepsilon > 0$ over the class of estimators being considered, where d is a metric on R^k. Of course, for any given class of estimators and any given metric, d, there is no guarantee that such an estimator exists.

Fortunately when the disturbance vector in (4.5.1) is assumed to be $E(n,\Sigma)$ with a density function of the form (3.2.7), we can prove the following:

Theorem 4.4. If the disturbance vector in (4.5.1) is distributed $E(n,\Sigma)$ with a density function of the form (3.2.7) then the GLS estimator (4.5.3) is the best linear estimator of β of the form (4.5.4) such that (4.5.5) holds, in the sense that for any $\varepsilon > 0$, any $k \times k$ symmetric positive semi-definite matrix, Γ, of rank $m \geq 1$, and for any other estimator, $\hat{\beta}$, of the form (4.5.4) for which (4.5.5) holds,

$$(4.5.7) \qquad \Pr\{(\hat{\beta}-\beta)'\Gamma(\hat{\beta}-\beta) < \varepsilon\} \leq \Pr\{(\tilde{\beta}-\beta)'\Gamma(\tilde{\beta}-\beta) < \varepsilon\}.$$

Proof: First note that if z is an $E(n,\Sigma)$ random vector with joint density function of the form (3.2.7) and if

$$E = \{v | v \in R^n,\ v'v < \varepsilon\}$$

for some scalar $\varepsilon > 0$, then

$$\Pr\{z \in E\} = \int_E |\Sigma|^{-\frac{1}{2}} \phi(z'\Sigma^{-1}z)\,dz$$

$$= \int_{E'} \phi(w'w)\,dw$$

by a simple change of variables, where

$$E' = \{v | v \in R^n,\ v'\Sigma v < \varepsilon\}$$

From (4.5.6) and property III,

$$\delta = \hat{\beta} - \beta$$

is $E(k, B\Sigma B')$. Because Γ is positive semi-definite of rank $m \geq 1$, there exists an $m \times k$ matrix, C, of rank m such that

$$\Gamma = C'C.$$

Then $C\delta$ is $E(m, CB\Sigma B'C')$ by property III.

Similarly,

$$\theta = \tilde{\beta} - \beta$$

is $E(k, (X'\Sigma^{-1}X)^{-1})$ and $C\theta$ is $E(m, C(X'\Sigma^{-1}X)^{-1}C')$.

Define the $k \times n$ matrix D by

$$D = B - (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}.$$

(4.5.5) implies $DX = 0$ and hence

$$CB\Sigma B'C' = C(X'\Sigma^{-1}X)^{-1}C' + CD\Sigma D'C'. \tag{4.5.8}$$

$$\begin{aligned} \Pr\{(\hat{\beta}-\beta)'\Gamma(\hat{\beta}-\beta) < \varepsilon\} &= \Pr\{\delta'C'C\delta < \varepsilon\} \\ &= \int_G \phi_m(w'w)\,dw, \end{aligned}$$

where $$G = \{v \mid v \in R^m,\ v'CB\Sigma B'C'v < \varepsilon\}$$

and $$\begin{aligned} \Pr\{(\tilde{\beta}-\beta)'\Gamma(\tilde{\beta}-\beta) < \varepsilon\} &= \Pr\{\theta'C'C\theta < \varepsilon\} \\ &= \int_{G'} \phi_m(w'w)\,dw, \end{aligned}$$

where $$G' = \{v \mid v \in R^m,\ v'C(X'\Sigma^{-1}X)^{-1}C'v < \varepsilon\}.$$

Because $CD\Sigma D'C'$ is positive semi-definite, (4.5.8) implies

$$G \subset G'$$

and (4.5.7) follows.

Corollary 4.4.1. If the disturbance vector in (4.5.1) is distributed $E(n,\Sigma)$ with a density function of the form (3.2.7), then for any k-dimensional, nonstochastic vector λ, such that $\lambda \neq 0$,

$$\Pr\{|\lambda'\hat{\beta}-\lambda'\beta| < \varepsilon\} \leqslant \Pr\{|\lambda'\tilde{\beta}-\lambda'\beta| < \varepsilon\} \tag{4.5.9}$$

for all real $\varepsilon > 0$ where $\tilde{\beta}$ is the GLS estimator (4.5.3) and $\hat{\beta}$ is any other linear estimator of the form (4.5.4) such that (4.5.5) holds.

Proof: Follows upon application of Theorem 4.4 with $\Gamma = \lambda\lambda'$.

For any component of β, β_i say, Corollary 4.4.1 implies that when the regression disturbances are elliptically symmetric, the estimate of β_i provided by $\tilde{\beta}_i$ has at least an equal probability of deviating by less than any given amount from β_i than the estimate provided by any other linear estimator of the form (4.5.4) such that (4.5.5) holds.

Clearly, optimality properties (4.5.7) and (4.5.9) hold for the OLS estimator within the class of linear estimators of the form (4.5.4) such that (4.5.5) holds when the regression disturbances are distributed $E(n,I_n)$ with joint density function of the form (3.2.1).

It is not immediately obvious that the optimality property, (4.5.7), is stronger than that upon which the Gauss-Markov theorem is based. Theorem 4.5 demonstrates that it is stronger.

Theorem 4.5. Let $\hat{\theta}$ and $\tilde{\theta}$ be any two unbiased estimators of θ, a k-dimensional vector of unknown parameters. If for any $\varepsilon > 0$ and any $k \times k$ symmetric positive semi-definite matrix, Γ, of rank $m \geqslant 1$, (4.5.10) $\Pr\{(\hat{\theta}-\theta)'\Gamma(\hat{\theta}-\theta) < \varepsilon\} \leqslant \Pr\{(\tilde{\theta}-\theta)'\Gamma(\tilde{\theta}-\theta) < \varepsilon\}$,

then if the second order moments of both $\hat{\theta}$ and $\tilde{\theta}$ exist,

(4.5.11) $$V(\hat{\theta}) = V(\tilde{\theta}) + D,$$

where $V(\cdot)$ denotes the covariance matrix operator and D is a $k \times k$ positive semi-definite matrix.

<u>Proof</u>: First note that if ω_1 and ω_2 are two positive random variables such that

$$\Pr(\omega_1<\varepsilon) \leqslant \Pr(\omega_2<\varepsilon)$$

for all $\varepsilon > 0$ then[16]

$$E(\omega_1) \geqslant E(\omega_2).$$

Let

$$D = V(\hat{\theta}) - V(\tilde{\theta}).$$

For any k-dimensional vector $\lambda \neq 0$, let $\Gamma = \lambda\lambda'$. Then (4.5.10) reduces to

$$\Pr\{(\lambda'\hat{\theta}-\lambda'\theta)^2 < \varepsilon\} \leqslant \Pr\{(\lambda'\tilde{\theta}-\lambda\theta)^2 < \varepsilon\}.$$

Because $\hat{\theta}$ and $\tilde{\theta}$ are unbiased and both $(\lambda'\hat{\theta}-\lambda'\theta)^2$ and $(\lambda'\tilde{\theta}-\lambda'\theta)^2$ are positive random variables,

$$\begin{aligned} V(\lambda'\hat{\theta}) &= E((\lambda'\hat{\theta}-\lambda'\theta)^2) \\ &\geqslant E((\lambda'\tilde{\theta}-\lambda'\theta)^2) \\ &= V(\lambda'\tilde{\theta}). \end{aligned}$$

---

16. See Parzen (1960, p.211).

Therefore,

$$\lambda' D\lambda = \lambda' V(\hat{\theta})\lambda - \lambda' V(\tilde{\theta})\lambda \geq 0$$

for all k-dimensional vectors λ such that $\lambda \neq 0$. Hence D is positive semi-definite as required.

Obviously (4.5.11) does not imply (4.5.10); hence Theorem 4.5 demonstrates that the optimality criterion of Theorem 4.4 is stronger than that of the Gauss-Markov theorem.

6. MAXIMUM LIKELIHOOD ESTIMATION

In this section we examine the maximum likelihood estimators of the unknown parameters in the linear model (4.5.1), when the disturbances are elliptically symmetric.

It is well known that when regression disturbances are distributed $N(0,\sigma^2\Sigma)$, then the GLS estimator $\tilde{\beta}$, given by (4.5.3), and

$$\tilde{\sigma}^2 = (y-X\tilde{\beta})'\Sigma^{-1}(y-X\tilde{\beta})/n \tag{4.6.1}$$

are the maximum likelihood estimators of β and σ^2 respectively. We shall show that the GLS estimator is the maximum likelihood estimator when the distribution of u belongs to a wide class of elliptically symmetric distributions. On the other hand, we shall find that when these elliptically symmetric disturbances have a finite covariance matrix $\sigma^2\Sigma_o$, $\tilde{\sigma}^2$ is the maximum likelihood estimator of σ^2 only for a comparatively small subclass of such disturbances.

Theorem 4.5. Suppose the disturbance vector, u, in the regression model (4.5.1) is an $E(n,\Sigma)$ random vector with joint density function of the form

$$(4.6.2) \qquad g(u) = |\Sigma|^{-\frac{1}{2}}\phi_n(u'\Sigma^{-1}u),$$

where ϕ_n is a non-increasing function on $[o,\infty)$. Then the GLS estimator (4.5.3) is the maximum likelihood estimator of β.

Proof: Let $\hat{\beta}$ denote the maximum likelihood estimator of β. The likelihood function of y is

$$L = |\Sigma|^{-\frac{1}{2}}\phi_n((y-X\beta)'\Sigma^{-1}(y-X\beta)).$$

Since ϕ_n is a nonincreasing function on $[o,\infty)$, L is maximized when its argument is minimized. The first order conditions for minimizing

$$(4.6.3) \qquad \pi = (y-X\beta)'\Sigma^{-1}(y-X\beta)$$

are

$$\frac{d\pi}{d\hat{\beta}} = -2X'\Sigma^{-1}(y-X\hat{\beta}) = 0,$$

which imply

$$(4.6.4) \qquad \hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y.$$

That

$$\frac{d^2\pi}{d\hat{\beta}^2} = 2X'\Sigma^{-1}X$$

is positive definite confirms the fact that the GLS estimator, (4.6.4), minimizes (4.6.3) and hence is the maximum likelihood estimator of β.

Corollary 4.5.1. If the joint density function of the disturbance vector u in (4.5.1) has the form

$$h(u) = \phi_n(u'u),$$

where ϕ_n is a non-increasing function on $[o,\infty)$, then the OLS estimator,

$$\text{(4.6.5)} \qquad b = (X'X)^{-1}X'y,$$

is the maximum likelihood estimator of β.

Note that a sufficient condition for ϕ_n in (4.6.2) to be non-increasing is that

$$h_n(x) = \phi_n(x'x)$$

be a marginal density function of an $E(n+2, I_{n+2})$ distribution with a joint density function of the form

$$h_{n+2}(z) = \phi_{n+2}(z'z),$$

where x and z are n-dimensional and (n+2)-dimensional vectors respectively. This follows because in this case, $\phi_n(x'x)$ can be written as

$$\begin{aligned}\phi_n(x'x) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_{n+2}(x'x+z_1^2+z_2^2)\,dz_1 dz_2 \\ &= \int_o^{2\pi}\int_o^{\infty} r\phi_{n+2}(x'x+r^2)\,dr d\theta \\ &= 2\pi\int_{(x'x)^{\frac{1}{2}}}^{\infty} (w^2-x'x)^{\frac{1}{2}}\phi_{n+2}(w^2)\,dw,\end{aligned}$$

which clearly is non-increasing in $x'x$. Therefore, in view of property VIII, the assumption that ϕ_n is a non-increasing function on $[o,\infty)$ is not particularly restrictive, especially in a time series context.

Theorem 4.6. If the disturbance vector, u, in the linear regression model (4.5.1) is an $E(n,\Sigma_o)$ random vector with covariance matrix

$$E(uu') = \sigma^2\Sigma_o, \qquad 0<\sigma^2<\infty,$$

and joint density function of the form

$$(4.6.6) \qquad g(u) = (\sigma^2)^{-n/2}|\Sigma_o|^{-\frac{1}{2}}\phi_o(u'\Sigma_o^{-1}u/\sigma^2),$$

where ϕ_o is a non-increasing function on $[o,\infty)$, differentiable over the range (o,∞), then a necessary condition for (4.6.1) to be the maximum likelihood estimator of σ^2 is that

$$(4.6.7) \qquad \dot{\phi}_o(n) = -\tfrac{1}{2}\phi_o(n),$$

where $\dot{\phi}_o(\cdot)$ denotes the first derivative of $\phi_o(\cdot)$.

Proof: Let $\hat{\beta}$ and $\hat{\sigma}^2$ denote the maximum likelihood estimators of β and σ^2 respectively. The likelihood function of y is

$$L = (\sigma^2)^{-n/2}|\Sigma_o|^{-\frac{1}{2}}\phi_o((y-X\beta)'\Sigma_o^{-1}(y-X\beta)/\sigma^2).$$

Theorem 4.5 implies that for all non-zero values of σ^2, L is maximized when $\beta = \tilde{\beta}$, hence

$$\hat{\beta} = \tilde{\beta},$$

where $\tilde{\beta}$ is the GLS estimator (4.5.3).

$$\frac{\partial L}{\partial\sigma^2} = -\frac{n}{2}(\sigma^2)^{-(n+2)/2}|\Sigma_o|^{-\frac{1}{2}}\phi_o((y-X\beta)'\Sigma_o^{-1}(y-X\beta)/\sigma^2)$$
$$- (\sigma^2)^{-(n+4)/2}(y-X\beta)'\Sigma_o^{-1}(y-X\beta)|\Sigma_o|^{-\frac{1}{2}}\dot{\phi}_o((y-X\beta)'\Sigma_o^{-1}(y-X\beta)/\sigma^2).$$

A necessary condition that $\tilde{\sigma}^2$ be the maximum likelihood estimator of σ^2 is that

$$\left.\frac{\partial L}{\partial \sigma^2}\right|_{\substack{\beta = \tilde{\beta} \\ \sigma^2 = \tilde{\sigma}^2}} = 0$$

or equivalently that

$$\dot{\phi}_o(n) = -\tfrac{1}{2}\phi_o(n),$$

as required.

Kelker (1970, p.422) found that a sufficient condition for ϕ_o to be differentiable for all values of its argument[17] is that

$$h_o(x) = \sigma^{-n}\phi_o(x'x/\sigma^2)$$

be a marginal density function of an $E(n+4, I_{n+4})$ distribution with an arbitrary joint density function when x is an n-dimensional vector. Therefore, the additional assumption that ϕ_o be differentiable over the range (o,∞) cannot be considered greatly restrictive.

Note that when u is distributed $N(0,\sigma^2\Sigma_o)$, ϕ_o has the form

$$\phi_o(\cdot) = (2\pi)^{-n/2}\exp(-\tfrac{1}{2}\cdot),$$

which satisfies (4.6.7). In order to better appreciate how restrictive condition (4.6.7) is, suppose u is an n-dimensional sample from a spherically invariant stochastic process. In this case, the joint density function of u can be written as

$$g(u) = \int_o^\infty (2\pi\sigma^2\tau^2)^{-n/2}|\Sigma_o|^{-\frac{1}{2}}\exp\{-u'\Sigma_o^{-1}u/(2\sigma^2\tau^2)\}dF(\tau),$$

where $F(\tau)$ is a distribution function supported on $[o,\infty)$. For

---

17. Except possibly at the origin.

(4.6.7) to hold, $F(\tau)$ must be such that

(4.6.8) $$\int_0^\infty \tau^{-n}(1-\tau^{-2})\exp({}^{-n}/2\tau^2)\,dF(\tau) = 0,$$

a condition which one would expect to be satisfied by comparatively few $F(\tau)$. Distribution functions, $F(\tau)$, which satisfy (4.6.8) only for a particular value of n, are of no statistical interest. The only $F(\tau)$ that have been found by the author to satisfy (4.6.8) for all values of n are of the form

$$F(\tau) = aU(\tau) + (1-a)U(\tau-1),$$

where

$$\begin{aligned} U(t) &= 0 \quad \text{for} \quad t \in (-\infty,0), \\ &= 1 \quad \text{for} \quad t \in (0,\infty) \end{aligned}$$ [18]

and $a \in [0,1]$. Such distribution functions result in u having a distribution which is a weighted sum of the degenerate distribution at $u = 0$ and the $N(0,\sigma^2\Sigma_0)$ distribution.

Clearly the maximum likelihood estimator of σ^2 depends on the form of ϕ_0 in (4.6.6). For example, when u has the multivariate Student-t distribution with joint density function

$$f(u) = p(\nu_0)(\tau^2)^{-n/2}\{\nu_0+u'u/\tau^2\}^{-(n-\nu_0)/2},$$

where $\tau > 0$ is unknown, $\nu_0 > 0$ is known and

$$p(\nu_0) = \nu_0^{\nu_0/2}\Gamma[(\nu_0+n)/2]/\pi^{n/2}\Gamma(\nu_0/2),$$

then if $\nu_0 > 2$,

---

18. $U(0)$ is undefined.

$$E(uu') = \nu_o \tau^2 / (\nu_o - 2) I_n$$
$$= \sigma^2 I_n$$

and Zellner (1976) has shown the maximum likelihood estimator of σ^2 to be

$$\hat{\sigma}^2 = \nu_o (y-Xb)'(y-Xb)/(\nu_o - 2),$$

where b is the OLS estimator, (4.6.5).

7. CONCLUSIONS

The main conclusion that follows from the results presented in this chapter is that the GLS estimator (and the OLS estimator as a special case) of β in the usual linear regression model has a number of desirable statistical properties when the disturbances are elliptically symmetric. Under reasonably lax conditions, it is weakly consistent, while under slightly stricter conditions it is strongly consistent. It is the best linear "unbiased" estimator with respect to an intuitively desirable and comparatively powerful optimality criterion. For a slightly restricted class of elliptically symmetric disturbances the GLS estimator of β is also the maximum likelihood estimator.

On the other hand, the GLS estimator is not asymptotically normal in general. When regression disturbances are independent but non-normal, under appropriate regularity conditions asymptotic normality may be used to justify the application of a number of statistical tests which are based on the assumption of normally distributed disturbances. As we shall find in the following chapter, this justification of the use of such statistical tests is superfluous in the case of elliptically symmetric disturbances. Therefore, with respect to the statistical properties of the GLS estimator considered in this chapter, it appears that little is lost if the assumption of normally distributed disturbances is weakened to one of elliptically symmetric disturbances.

Finally we close this chapter with the conjecture that for disturbances with finite covariance matrix

$$E(u(n)u'(n)) = \Sigma(n)$$

such that (4.2.6) and (4.2.7) hold, the widest class of disturbance distributions for which (4.2.5) is a necessary and sufficient condition for strong consistency of the GLS estimator is the class of $\mathcal{E}(n,\Sigma(n))$ distributions with finite second order moments.

CHAPTER 5

SMALL SAMPLE PROPERTIES OF TESTS AND ESTIMATORS WHEN THE REGRESSION DISTURBANCES ARE ELLIPTICALLY SYMMETRIC

1. INTRODUCTION

This chapter investigates the small sample properties of statistical tests and some better known estimators associated with the linear regression model when the disturbance vector is assumed to follow an elliptically symmetric distribution. Although the principal results of this chapter are established in the more general context of a non-linear regression model, the only applications discussed are those relating to the special case of the linear model.

One of the two cornerstones of this chapter is the result established in Section 2; that with respect to the general non-linear regression model, any statistic which is invariant to the scale of the disturbances has the same distribution for all $E_o(n,\Sigma)$ distributions taken by the disturbance vector. Section 3 is devoted to the implications of this result for the significance level and hence the validity of statistical tests associated with the linear regression model with $E_o(n,\Sigma)$ disturbances. The power of tests for serial correlation and heteroscedasticity in regression disturbances which take an appropriate elliptically symmetric distribution are investigated in Section 4. Optimal power properties of such tests known to hold under normality are found to hold also for elliptically symmetric disturbances.

The second cornerstone is the result presented in Section 5 concerning the distribution of non-linear regression statistics. It is found that whenever the disturbance vector takes a suitably well-behaved $E(n,\Sigma)$ distribution, the distribution of any such statistic can be viewed as a weighted average of the statistic's distributions for different values of σ under the assumption of $N(0,\sigma^2\Sigma)$ disturbances. This allows us to discuss the validity of tests whose statistics are not invariant to the scale of the disturbances in both the usual linear regression model and the lagged dependent variable regression model. It also enables statements to be made about the power properties of tests other than those already considered in Section 4.

The implications of this second major result for the distributions of estimators of linear regression coefficients and nuisance parameters, is the subject of Section 6. The final section contains some concluding remarks.

2. A THEORETICAL RESULT

In this section we prove that for a given characteristic matrix, Σ, the distributions of a wide class of non-linear regression statistics are invariant to the particular type of elliptically symmetric distribution the regression disturbances follow. This is an important result because it establishes a wide class of statistics whose distributions are unchanged when the usual assumption of normality is replaced by the more general assumption of elliptical symmetry.

Consider the non-linear regression model,

$$y = f(\theta,v), \tag{5.2.1}$$

where y is an observable, n-dimensional random vector, f is a known, Borel measurable, vector function such that

$$f: R^k \times R^m \to R^n,$$

θ is a k-dimensional vector of parameters which may be unknown, and v is an unobservable, m-dimensional random vector.[1] Let σ^2 be any positive scalar and Σ be any $m \times m$ positive definite matrix.

<u>Theorem 5.1</u>. With respect to the non-linear regression model (5.2.1), any statistic, which is invariant to the scale of v, has the same distribution when v follows the $N(0,\sigma^2\Sigma)$ distribution as it does for v taking any other $E_o(m,\Sigma)$ distribution.

<u>Proof</u>: Denote the statistic by $g(y)$. Define the function,

$$\bar{g}: R^m \to R,$$

by

$$\bar{g}(v) = g(f(\theta,v)),$$

and let G be the group of transformations of the form

$$v \to \lambda v, \tag{5.2.2}$$

where λ is any positive scalar.

Consider the statistic

$$\eta(v) = v/(v'\Sigma^{-1}v)^{\frac{1}{2}}.$$

---

1. We have allowed the dimension of y and v to differ in order to be completely general. In most applications the values taken by m and n will be the same.

Clearly $\eta(v)$ is invariant to transformations belonging to G. If $v_{(1)}$ and $v_{(2)}$ are two m-dimensional vectors such that

$$\eta(v_{(1)}) = \eta(v_{(2)}),$$

then $v_{(1)}$ and $v_{(2)}$ are related by a transformation of the form of (5.2.2). Therefore, $\eta(v)$ is a maximal invariant.[2] Because $\bar{g}$ is also invariant to transformations belonging to G, it can be written as a function of $\eta(v)$, i.e.,

$$\begin{aligned} g(y) &= \bar{g}(v) \\ &= g^*(v/(v'\Sigma^{-1}v)^{\frac{1}{2}}), \end{aligned}$$

say. When v takes an $E_o(m,\Sigma)$ distribution, $\Sigma^{-\frac{1}{2}}v$ is $E_o(m,I_m)$ and by property X,

$$\Sigma^{-\frac{1}{2}}v/(v'\Sigma^{-1}v)^{\frac{1}{2}},$$

is uniformly distributed on the surface of the n-dimensional unit sphere. Thus

$$v/(v'\Sigma^{-1}v)^{\frac{1}{2}},$$

and hence $g(y)$, have the same distributions for any $E_o(m,\Sigma)$ distribution followed by v, including $N(0,\sigma^2\Sigma)$.

In a nutshell, Theorem 5.1 establishes a remarkable property of elliptically symmetric distributions that appears to have gone unnoticed in the literature. The property is that for any given characteristic matrix Σ, any scale invariant function[3] of the random vector v, has

---

2. For the definition and the derivation of some of the properties of a maximal invariant see Lehmann (1959, p.215).

3. $\bar{g}(\cdot)$ in the proof of Theorem 5.1.

the same distribution for all $E_o(m,\Sigma)$ distributions v may follow. In Theorem 5.1, the scale invariant function has been split into two functions, the non-linear regression model (5.2.1) and the statistic $g(y)$.

Theorem 5.1 is a very versatile result. We shall restrict our attention to the following two special cases of (5.2.1):

(i) The usual linear regression model

(5.2.3) $$y = X\beta + u,$$

where X is an $n \times k$, nonstochastic[4] matrix, β is a k-dimensional vector of parameters and u is an n-dimensional disturbance vector. Throughout this chapter, H_k will be used to denote this model. Two special cases of H_k worth bearing in mind are

$$y = u$$

and

(5.2.4) $$y = \mu\ell + u,$$

where μ is a scalar and ℓ is the n-dimensional vector $\ell = (1,1,\ldots 1)'$.

(ii) The linear regression model with a lagged dependent variable as a regressor,

(5.2.5) $$y = \alpha y_{-1} + X\beta + u,$$

---

4. Alternatively the elements of X may be assumed to be distributed independently of u. In this case the expectation operator is the expectation operator conditional on X. Also note that no assumption has been made about the rank of X.

where α is a scalar, $y_{-1} = (y_o, y_1, \ldots y_{n-1})'$ and X, β and u are as in H_k. H_k^{-1} will be used to denote this model with the additional assumptions that for some non-negative integer p,

(a) $\tilde{u}' = (u_{-p+1}, u_{-p+2}, \ldots, u_o, u')$ is independent of y_{-p} and

(b) $\tilde{u}$ has an elliptically symmetric distribution.

These two assumptions allow Theorem 5.1 to be applied to (5.2.5). Let Π be the $n \times n$ matrix

$$\Pi = \begin{bmatrix} 1 & 0 & 0 & & 0 \\ \alpha & 1 & 0 & & 0 \\ \alpha^2 & \alpha & 1 & & 0 \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \alpha^{n-1} & \alpha^{n-2} & \alpha^{n-3} & \cdots & 1 \end{bmatrix}$$

and let ζ be the n-dimensional vector

$$\zeta = \begin{bmatrix} \alpha \\ \alpha^2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha^n \end{bmatrix}.$$

Then (5.2.5) can be rewritten as

$$y = y_o \zeta + \Pi X \beta + \Pi u,$$

i.e. y can be expressed as a function of the stochastic elements y_0 and u. By substituting in turn for y_0, y_{-1}, ... and y_{-p+1}, y can also be expressed as a function of y_{-p} and $\tilde{u}$. Thus assumptions (a) and (b) allow Theorem 5.1 to be applied with

$m = n + p.$[5]

H_k^{-1} is a special case of the dynamic linear model

$$y = \alpha_1 y_{(-1)} + \alpha_2 y_{(-2)} + \ldots + \alpha_j y_{(-j)} + X\beta + u,$$

where

$$y_{(-i)} = (y_{1-i}, y_{2-i}, \ldots, y_{n-i})', \qquad i=1,\ldots,j.$$

If for some non-negative interger p,

(a) $\tilde{u}$ is independent of $y_{-p}, y_{-(p+1)}, \ldots y_{-(p+j-1)}$ and

(b) $\tilde{u}$ is elliptically symmetric, then clearly this model also fulfills the requirements of Theorem 5.1 for $m = n + p$.

3. IMPLICATIONS FOR THE VALIDITY OF STATISTICAL TESTS

An immediate implication of Theorem 5.1 is that any statistical test whose null hypothesis is H_k and whose test statistic is invariant to the scale of the disturbances, has, for a given regression model (5.2.3), the same size whether the disturbance vector is distributed $N(0,\sigma^2\Sigma)$ or follows any other $E_o(n,\Sigma)$ distribution. Hence such statistical tests are equally as valid, either as approximate tests or exact tests, which ever the case may be, when u is $E_o(n,\Sigma)$, as they are under the usual assumption that u is $N(0,\sigma^2\Sigma)$. Since almost always under this latter assumption, σ^2 is assumed to be unknown, the great majority of small sample tests and a number of

---

5. Theorem 5.1 implies that any statistic, which is invariant to the scale of $\tilde{u}$ in (5.2.5), has the same conditional distribution with respect to the value taken by y_{-p} for all $E_o(n+p,\Sigma)$ distributions $\tilde{u}$ may follow. Independence of y_{-p} and $\tilde{u}$ implies the unconditional distribution of the statistic is also invariant to the particular type of $E_o(n+p,\Sigma)$ distribution $\tilde{u}$ follows.

asymptotic tests are based on test statistics which are invariant to the positive values taken by σ^2.

A list of such statistical tests would include the following:

(i) All variations of the classical F-test of fixed linear restrictions on the coefficients of H_k including the test of individual coefficient values based on the Student's t distributed statistic and the tests of equality between sets of coefficients in two linear regressions recently reviewed by Fisher (1970).

(ii) The Durbin-Watson (1950, 1951) and the Berenblut-Webb (1973) bounds tests for first-order autocorrelation in regression disturbances using either the lower bound, upper bound, true significance point or any approximation[6] to the true significance point as the critical value.

(iii) Generalizations of the DW (Durbin-Watson) bounds test or the Berenblut-Webb bounds test proposed as tests for higher order autocorrelation in regression disturbances by Wallis (1972), Schmidt (1972), Vinod (1973) and Webb (1973) using either the lower bound, upper bound, true significance point or any approximation to the true significance point as the critical value.

(iv) Tests for serial correlation in regression disturbances suggested by Anderson (1948), Anderson and Anderson (1950),

---

6. Alternative methods of approximating the true critical value of the DW test statistic have been reviewed by Durbin and Watson (1971) and Harrison (1972).

Hannan (1955, 1957), Griliches et al. (1962), Theil (1965, 1968), Durbin (1967, 1969, 1970b), Hannan and Terrell (1968), Geary (1970), Kadiyala (1970), Abrahamse and Koerts (1971), Thomas and Wallis (1971), Dubbelman et al. (1972), Cliff and Ord (1973), Dent (1973), Phillips and Harvey (1974), Sims (1975), Fraser et al. (1976), Bartels and Hordijk (1977), Savin and White (1978) and Dent and Styan (1978) as well as tests based on von Neumann's ratio using LUSH (Linear Unbiased with Scalar covariance matrix formed using Householder transformations) residuals which may be obtained by Golub's (1965) procedure.

(v) Approximate non-parametric tests for serial correlation in regression disturbances that were originally constructed as tests for the independence of a zero-mean random sample but which are often applied to regression residuals, the best known perhaps being the Runs test proposed by Wald and Wolfowitz (1940) for which Swed and Eisenhart (1943) tabulated cumulative probabilities.

(vi) Tests for heteroscedastic regression disturbances proposed by Goldfeld and Quandt (1965), Putter (1967), Glejser (1969), Ramsey (1969), Heyadat and Robson (1970), Theil (1971, p.214) and Harvey and Phillips (1974). Also asymptotic tests for multiplicative heteroscedasticity using the likelihood ratio, Wald's and the Lagrange multiplier approaches as outlined by Harvey (1976) and Godfrey (1978b).

(vii) Further tests for misspecification of H_k not included in (i) above, such as those outlined by Ramsey (1969), Farley and Hinich (1970), Theil (1971, p.222), Brown et al. (1975), Ramsey and Schmidt (1976), Harvey and Collier (1977), Jayatissa (1977) and Lyons and Proctor (1977).

(viii) Tests for outlying observations in H_k such as those discussed by Daniel (1960), Ferguson (1961), Stefansky (1971, 1972), Ellenberg (1973, 1976), Goldsmith and Boddy (1973), Tietjen et al. (1973), Williams (1973), Lund (1975) and Prescott (1975).

(ix) Tests of the normality of regression disturbances suggested by Putter (1967), Koerts and Abrahamse (1969, p.125), Louter and Koerts (1970), Huang and Bolch (1974) and Mukantseva (1977).

Note that it is not realistic to describe members of this latter group as tests for normality. They are in fact tests for spherical symmetry. Each of these tests was originally designed to test the null hypothesis that u is distributed $N(0,\sigma^2 I_n)$, where $\sigma^2 > 0$ is unknown. Were the alternative hypothesis to be that u is $E_o(n,I_n)$, then the probability of rejecting normality would be the same under both the null and the alternative hypotheses. The same result could have been achieved by the randomized decision rule which rejects normality with probability α, where α is the desired size of the test. This, together with the fact that the sizes of such tests remain unaltered when the null hypothesis is widened to allow u to take any other $E_o(n,I_n)$ distribution, means these tests are, in reality, tests for spherical symmetry and should be regarded as such.

The above list of tests is far from complete. There is a wealth of statistical tests proposed in the literature that are based on (5.2.4) with the value taken by μ either being unknown or assumed to take a specific value. For example, the size of any statistical test whose null hypothesis is that the observable, n-dimensional random vector y, is $N(0,\sigma^2\Sigma)$ and whose test statistic is invariant to the scale of y, is independent of the particular type of $\mathcal{E}_o(n,\Sigma)$ distribution y may follow. Perhaps the most prominent subclass of such tests are those which are designed to test y for normality. In view of the comments made above, such tests should be regarded as tests for elliptical symmetry.

Another example is the use of the von Neumann (1941) ratio together with the critical values tabulated by Hart (1942) as a test of the hypothesis that y is generated by (5.2.4) with u distributed $N(0,\sigma^2 I_n)$, where μ and σ^2 are both unknown. Theorem 5.1 implies that it is also a valid test of the hypothesis that $y - \mu\ell$ is distributed $\mathcal{E}_o(n,I_n)$.

A further implication of Theorem 5.1 under H_k is that confidence intervals for individual regression coefficients or linear combinations of coefficients, constructed assuming $N(0,\sigma^2\Sigma)$ disturbances with σ^2 unknown and based on the familiar Student's t distributed statistic, remain unchanged when the disturbances are assumed to take any other $\mathcal{E}_o(n,\Sigma)$ distribution. Similarly, joint confidence regions on subsets of β_i, $i = 1, \ldots, k$, such as the usual minimum volume confidence ellipsoids based on F statistics, Bonferroni t-intervals[7] or Scheffé's (1959, p.58) S-method confidence regions, are equally as valid under the wider elliptical symmetry assumption as they are under their original normality assumption. Note that for a given set of

---

7. For a description of these confidence regions see Seber (1977, p.126).

observations on H_k, exactly the same confidence region results whether the disturbances are known to take a normal distribution for which all moments exist, or an elliptically symmetric Cauchy distribution for which no moments exist.

Theorem 5.1 also has implications for plots of residuals which are often used for investigating departures from the assumed model, H_k, in a less formal manner than hypothesis testing.

With respect to H_k with X of rank $k < n$ and u distributed $N(0,\sigma^2\Sigma)$, let $\tilde{u}$ denote the vector of GLS residuals,

$$\begin{aligned}\tilde{u} &= y - X\tilde{\beta} \\ &= [I_n - X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}]y \\ &= [I_n - X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}]u,\end{aligned}$$

where $\tilde{\beta}$ is the GLS estimator (4.5.3). Any function of these residuals that is invariant to the scale of u, satisfies the conditions of Theorem 5.1. In particular, normed residuals[8] given by

$$\tilde{u}_i/(\tilde{u}'\Sigma^{-1}\tilde{u})^{\frac{1}{2}}, \qquad i=1,\ldots,n; \tag{5.3.1}$$

scaled residuals[9] given by

$$\tilde{u}_i/\{s^2(n-k)/n\}^{\frac{1}{2}}, \qquad i=1,\ldots,n;$$

and "Studentized" residuals[10] given by

---

8. See Andrews (1971).

9. See Daniel and Wood (1971, p.28) or Seber (1977, p.163).

10. See Behnken and Draper (1972).

$$\tilde{u}_i/\{s^2(1-m_{ii})\}^{\frac{1}{2}}, \qquad i=1,\ldots,n,$$

where $\tilde{u}_i$ is the i^{th} component of $\tilde{u}$,

$$s^2 = \tilde{u}'\Sigma^{-1}\tilde{u}/(n-k)$$

and m_{ii} is the i^{th} diagonal element of

$$X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1},$$

all satisfy the conditions of Theorem 5.1 and hence their joint distributions are invariant to the particular $E_o(n,\Sigma)$ distribution u follows. Therefore, the array of statistical techniques that have been developed assuming normally distributed disturbances, and that are used to analyze various plots of these residuals, remain equally valid techniques when the joint distribution of the disturbances is any other $E_o(n,\Sigma)$ distribution.[11]

In this section, there is no example of an application of Theorem 5.1 to H_k^{-1}. In contrast to the case for H_k, there does not appear to be a test based on a statistic which is invariant to the scale of u as required for the application of Theorem 5.1. In general, tests concerning H_k^{-1} are constructed assuming normally distributed disturbances with an unknown scale parameter. Almost without exception such tests are asymptotic tests with small sample null distributions which vary with σ^2 but which tend to an asymptotic null distribution that is independent of σ^2. The validity of such tests for a class of $E(n,\Sigma)$ disturbances is discussed in Section 5.

---

11. For a recent review of such techniques see Seber (1977, pp. 162-172).

The main thrust of this section is that a large number of important statistical tests connected with the linear regression model, H_k, are as equally valid for elliptically symmetric disturbances as they are for normally distributed disturbances. Logically, the next question we should ask is whether or not such tests have useful power properties for elliptically symmetric disturbances. The following two sections attempt to answer this question.

4. POWER PROPERTIES OF TESTS FOR SERIAL CORRELATION AND HETEROSCEDASTICITY

For a number of statistical tests, the above question can be answered with the aid of Theorem 5.1. The following corollary shows that the power functions of a range of tests are invariant to the particular type of elliptically symmetric distribution the disturbances follow.

Corollary 5.1.1. Let Ω denote the class of $m \times m$ positive definite matrices and let Ω_0 and Ω_1 be any two disjoint, non-null subsets of Ω. With respect to the non-linear regression model (5.2.1), any test of the null hypothesis,

$$H_0 : v \sim E_0(m,\Sigma_0), \qquad \Sigma_0 \in \Omega_0,$$

against the alternative hypothesis,

$$H_a : v \sim E_0(m,\Sigma_1), \qquad \Sigma_1 \in \Omega_1,$$

with a test statistic which is invariant to the scale of v, has for any fixed $\Sigma_0 \in \Omega_0$, the same size for all $E_0(m,\Sigma_0)$ distributions followed by v and, for any fixed $\Sigma_1 \in \Omega_1$, the same power for all $E_0(m,\Sigma_1)$ distributions taken by v.

One of the major consequences of Corollary 5.1.1 is that all tests for serial correlation and heteroscedasticity in the disturbances of H_k, based on test statistics which are invariant to the scale of u, have the same power function whether the disturbances are normally distributed or follow an appropriate elliptically symmetric distribution. That is, the tests listed in (ii), (iii), (iv), (v) and (vi) of the previous section[12] have the same power for normally distributed disturbances as they do when the disturbances follow any other elliptically symmetric distribution with the same characteristic matrix.

Hence studies which have compared the power functions of these tests, either calculated exactly or estimated by Monte Carlo methods for normal disturbances, are equally as valid for disturbances taking any other appropriate elliptically symmetric distribution.

There are a large number of these studies reported in the literature. In the case of tests for serially correlated disturbances, any list would include works authored by Durbin (1967), Koerts and Abrahamse (1968, 1969), Abrahamse and Louter (1971), Dubbelman (1972, 1978), Habibagahi and Pratschke (1972), Berenblut and Webb (1973) Blattberg (1973), Cleur (1973), Cliff and Ord (1973, Chapter 7), Dent (1973), Phillips and Harvey (1974), Harrison (1975), L'Esperance and Taylor (1975), Schmidt and Guilkey (1975), Tillman (1975), Fraser et al. (1976), Smith (1976, 1977a, 1977b), Bartels and Hordijk (1977), Dent and Styan (1978), King and Giles (1978), and Dent and Cassing

---

12. The tests listed in (viii) can also be included if the alternative hypothesis is that the outlying observation has a disturbance term with larger variance than the remaining disturbances.

(1978). The power functions of tests for heteroscedastic normally distributed disturbances have been investigated by Goldfeld and Quandt (1965, 1972), Glejser (1969), Ramsey and Gilbert (1972) and Harvey and Phillips (1974).

Harrison and McCabe (1975) and Epps and Epps (1977) separately studied the robustness of selected tests for serial correlation and heteroscedasticity when both first-order autocorrelation and heteroscedasticity occur simultaneously in normally distributed regression disturbances. Conclusions drawn from these studies are equally as valid under an appropriate elliptical symmetry assumption as they are under the original normality assumption. Similarly, conclusions made by Toyoda (1974) and Schmidt and Sickles (1977) concerning the true significance level of the Chow (1960) test when the regression disturbances are heteroscedastic, are invariant to the particular type of elliptically symmetric distribution the disturbances are assumed to follow.

We now turn our attention to statistical tests known to possess optimal power properties under the usual normality assumption. The question we shall attempt to answer is whether such optimal power properties still hold under an equivalent elliptical symmetry assumption. Recent work by Kariya and Eaton (1977) and Kariya (1977) have provided some answers to this question for selected statistical tests, but only for elliptically symmetric distributions with a joint density function of the form (3.2.7), where ϕ is either a non-increasing function or a convex, non-increasing function on $[o,\infty)$. The following corollary of Theorem 5.1 allows almost all of these results to be generalized to $E_o(n,\Sigma)$ distributions. It also allows us to show that optimal power properties of tests for autocorrelation and heteroscedasticity, known to hold for normal disturbances, also hold, within the class of tests invariant to the scale of the disturbances, under the appropriate elliptical symmetry assumption.

Corollary 5.1.2. Let Ω denote the class of $m \times m$ positive definite matrices and let Ω_0 and Ω_1 be any two disjoint, non-null subsets of Ω. Suppose C_I denotes the class of statistical tests which, with respect to the non-linear regression model (5.2.1), are tests of the null hypothesis,

$$H_0 : v \sim E_0(m,\Sigma_0), \qquad \Sigma_0 \in \Omega_0,$$

against the alternative hypothesis,

$$H_a : v \sim E_0(m,\Sigma_1), \qquad \Sigma_1 \in \Omega_1,$$

and which are invariant to the scale of v. Consider a statistical test belonging to C_I. Suppose in the special case of normally distributed disturbances under both H_0 and H_a, the test is known to possess a particular optimal power property within a given class of tests denoted T.

(i) If $T \subset C_I$, the test has the same optimal power property within T, when testing H_0 against H_a.

(ii) If $C_I \subset T$, the test has the same optimal power property within C_I, when testing H_0 against H_a.

(iii) If $C_I \not\subset T$ and $T \not\subset C_I$, the test has the same optimal power property within $C_I \cap T$, when testing H_0 against H_a.

Note that the problem of testing H_0 against H_a is invariant to transformations of the form

$$y \to \alpha y, \tag{5.4.1}$$

where α is a positive scalar. It is also invariant to transformations of the form

$$y \to \alpha y + X\gamma, \tag{5.4.2}$$

where α is a positive scalar and γ is a k-dimensional vector. Hence C_I contains the class of tests invariant to transformations of the form (5.4.1) and both these classes contain those tests invariant to transformations of the form (5.4.2).

For the remainder of this section, let $E_\phi(n,\Sigma)$ denote an $E(n,\Sigma)$ distribution with a density function of the form of (3.2.7), where ϕ is such that (3.2.2) and (3.2.3) hold. Clearly an $E_\phi(n,\Sigma)$ distribution is a special case of an $E_o(n,\Sigma)$ distribution.

Kariya and Eaton (1977) considered the problem of testing the null hypothesis that a random vector x is $E_\phi(n,I_n)$, against the alternative that it is $E_\phi(n,\Sigma_1)$, with ϕ restricted to the class of non-increasing functions on $[o,\infty)$ and where Σ_1 is a fixed matrix. They showed that the test which rejects the null hypothesis for small values of

$$r = x'\Sigma_1^{-1}x/x'x, \tag{5.4.3}$$

is UMP (Uniformly Most Powerful) and that the null distribution of r is the same as that when x is $N(0,\sigma^2 I_n)$. Corollaries 5.1.1 and 5.1.2 allow this result to be generalized as follows:

<u>Corollary 5.1.3</u>. Let Σ_1 be a fixed, positive definite, $n \times n$ matrix such that $\Sigma_1 \neq \sigma^2 I_n$. With respect to the problem of testing the null hypothesis,

$$H_0 : \text{the random vector } x \text{ is } E_o(n, I_n),$$

against the alternative hypothesis,

$$H_a : x \text{ is } E_o(n, \Sigma_1),$$

the test which rejects H_0 for small values of (5.4.3) is UMPI, where invariance is with respect to transformations of the form (5.4.1). Under H_0, the distribution of the test statistic, (5.4.3), is the same as that when x is $N(0, \sigma^2 I_n)$, while under H_a it has the same distribution as when x is $N(0, \sigma^2 \Sigma_1)$.

Kariya and Eaton also showed that their result could be extended to the following situations when Σ_1 is not fixed:

(i) $\Sigma_1 = \sigma^2 \bar{\Sigma}_1$, $\bar{\Sigma}_1$ known,

(ii) $\Sigma_1 = \lambda_1(I_n - M) + \lambda_2 M$, $M^2 = M$, M known, where $\lambda_1 > \lambda_2 > 0$ (or $\lambda_2 > \lambda_1 > 0$), and

(iii) $\Sigma_1^{-1} = \lambda_1 I_n + \lambda_2 A$, A a known $n \times n$ matrix such that Σ_1^{-1} is positive definite and $\lambda_1 > 0$.

In each case, a UMP test was found that does not depend on the unknown parameters. Corollary 5.1.2 implies that these tests are also UMPI against the alternative hypothesis that x is $E_o(n, \Sigma_1)$, where invariance is with respect to transformations of the form (5.4.1).

Kariya (1977) studied the problem of testing the null hypothesis that the random vector x is $E_\phi(n, I_n)$, against the alternative that it is $E_\phi(n,\gamma\Sigma_1(\lambda))$, where $\gamma > 0$ and $\Sigma_1(\lambda)$ has the form

(5.4.4) $$\Sigma_1(\lambda) = [I_n + \lambda A]^{-1}$$

with $\lambda \in \Lambda = \{\lambda \in R | \Sigma_1^{-1}(\lambda)$ positive definite$\}$ and λ unknown. He found that when $A \neq \sigma^2 I_n$ and $A \neq 0$, the test statistic

(5.4.5) $$s = x'Ax/x'x$$

with a suitable critical region of the form $s < c_1$ and $s > c_2$, provides a UMPU (UMP Unbiased) test against the two-sided alternative that x is $E_\phi(n,\gamma\Sigma_1(\lambda))$, $\lambda \neq 0$, with ϕ restricted to the class of convex, non-increasing functions on $[o,\infty)$.

By applying Corollaries 5.1.1 and 5.1.2, this result can be generalized to the following:

<u>Corollary 5.1.4</u>. Let $\Sigma_1(\lambda)$ be a positive definite, $n \times n$ matrix of the form (5.4.4) such that $A \neq \sigma^2 I_n$ and $A \neq 0$. With respect to testing the null hypothesis,

$$H_0 : \text{the random vector } x \text{ is } E_o(n,I_n),$$

against the alternative hypothesis,

$$H_a : x \text{ is } E_o(n,\Sigma_1(\lambda)), \quad \lambda \neq 0,$$

the test statistic (5.4.5) with an appropriate critical region of the form $s < c_1$ or $s > c_2$ is a UMPUI test, where invariance is with respect to transformations of the form (5.4.1). Under H_0, the distribution of s is the same as that when x is $N(0,\sigma^2 I_n)$, while for each value of λ under H_a, s has the same distribution as when x is $N(0,\sigma^2\Sigma_1(\lambda))$.

Kariya applied his result, as well as the Kariya and Eaton result for the one-sided test, to the linear regression model, H_k, with the design matrix, X, assumed to be of full column rank.[13] In particular, Kariya considered the problem of testing the null hypothesis that the disturbance vector, u, is $E_\phi(n,I_n)$, against the alternative hypothesis that it takes an $E_\phi(n,\Sigma_1(\lambda))$ distribution, where $\Sigma_1(\lambda)$ is of the form of (5.4.4) and either $\lambda > 0$ or $\lambda \neq 0$. Under the one-sided alternative hypothesis, ϕ is assumed to be non-increasing on $[o,\infty)$, while for the two-sided alternative hypothesis it is assumed to be non-increasing and convex on $[o,\infty)$. Kariya showed that when the column space of the design matrix, X, is spanned by some k characteristic vectors of A, $A \neq 0$ and $A \neq \sigma^2 I_n$, the test statistic,

$$s^* = y'MAMy/y'My, \tag{5.4.6}$$

where

$$M = I_n - X(X'X)^{-1}X',$$

with critical region $s^* < c_0$ yields a UMPI[14] test in the case of the one-sided alternative hypothesis, while in the two-sided case, (5.4.6) with an appropriate critical region of the form $s^* < c_1$ or $s^* > c_2$ was shown to be UMPUI.

Again these results can be generalized to the following, using Corollaries 5.1.1 and 5.1.2:

---

13. Whenever the regression model H_k is referred to in the remainder of this section, the design matrix will be assumed to be of full column rank.

14. Here, and for the remainder of this section, invariance is with respect to transformations of the form (5.4.2).

<u>Corollary 5.1.5</u>. Suppose the linear regression model H_k holds and $u \sim E_o(n,\Sigma_1(\lambda))$, where $\Sigma_1(\lambda)$ is of the form (5.4.4) with $A \neq \sigma^2 I_n$ and $A \neq 0$. If the column space of the design matrix, X, is spanned by some k characteristic vectors of A, the test statistic (5.4.6) with the critical region $s^* < c_0$ is a UMPI test of the null hypothesis,

$$H_0: \lambda = 0,$$

against the alternative hypothesis,

$$H_a: \lambda > 0,$$

while with an appropriate critical region of the form $s^* < c_1$ or $s^* > c_2$, it is a UMPUI test of H_0 against the two-sided alternative,

$$H_a': \lambda \neq 0.$$

For any given value of λ, the distribution of s^* is the same for all $E_o(n,\Sigma_1(\lambda))$ distributions taken by u.

In the remainder of this section, the above results are applied to some of the better known tests for serial correlation in the disturbances of H_k, such as the DW test, the Wallis test and the Berenblut-Webb test.

Suppose under H_k, the components of u are generated by the stationary, first-order, autoregressive scheme,

$$u_t = \rho u_{t-1} + e_t, \quad |\rho| < 1, \quad t=1,2,\ldots, \tag{5.4.7}$$

where $e = (e_1,\ldots,e_n)'$ is $E_o(n,I_n)$. Property III implies u is distributed $E_o(n,\Sigma(\rho))$, where $\Sigma(\rho)$ is of the form

$$\Sigma(\rho) = [(1-\rho)^2 I_n + \rho A_1 + \rho(1-\rho)C_1]^{-1}, \tag{5.4.8}$$

with A_1 and C_1 given by (4.2.29) and (4.2.30), respectively.

Durbin and Watson (1950) approximated (5.4.8) with

$$\bar{\Sigma}(\rho) = [(1-\rho)^2 I_n + \rho A_1]^{-1}$$

and assumed normally distributed disturbances. This enabled them to show, using a result due to Anderson (1948), that their test statistic,

$$d_1 = y'MA_1My/y'My, \tag{5.4.9}$$

with critical region $d_1 < c_0$, is an approximately UMP similar test of $H_0: \rho = 0$ against $H_a: \rho > 0$, when the column space of X is spanned by k characteristic vectors of A_1. Under these conditions, Durbin and Watson (1971) also found their test to be UMPI. Kariya (1977) showed that the one-sided DW test is UMPI when e is $E_\phi(n,I_n)$ with ϕ restricted to the class of non-increasing functions on $[o,\infty)$ and the column space of X is spanned by k characteristic vectors of A_1. With this same restriction on X, he also concluded that the two-sided unbiased DW test is an approximately UMPUI test of $H_0: \rho = 0$ against $H_a': \rho \neq 0$, when ϕ is a non-increasing and convex function on $[o,\infty)$. Corollaries 5.1.1 and 5.1.2 allow these results to be extended to the following:

<u>Corollary 5.1.6</u>. Suppose the linear regression model H_k holds with $u \sim E_o(n,\Sigma(\rho))$, where $\Sigma(\rho)$ is given by (5.4.8). If the column space of X is spanned by some k characteristic vectors of A_1, the DW statistic, (5.4.9), with critical region $d_1 < c_0$, is an approximately

UMPI test of the null hypothesis,

$$H_0 : \rho = 0,$$

against the alternative

$$H_a : \rho > 0,$$

while with an appropriate critical region of the form $d_1 < c_1$ or $d_1 > c_2$, it is an approximately UMPUI test of H_0 against the two-sided alternative,

$$H_a' : \rho \neq 0.$$

For any given value of ρ, the distribution of d_1 is the same for all $E_o(n,\Sigma(\rho))$ distributions taken by u including the $N(0,\sigma^2\Sigma(\rho))$ distribution.

Assuming e to be $N(0,\sigma^2 I_n)$, Durbin and Watson (1971) showed that their one-sided test is approximately locally UMPI in the neighbourhood of $\rho = 0$. The application of Corollary 5.1.2 to this result yields:

<u>Corollary 5.1.7</u>. Suppose the linear regression model, H_k, holds with $u \sim E_o(n,\Sigma(\rho))$, where $\Sigma(\rho)$ is given by (5.4.8). The DW statistic, (5.4.9), with critical region $d_1 < c_0$, is an approximately locally UMPI test of the null hypothesis,

$$H_0 : \rho = 0,$$

against the alternative hypothesis,

$$H_a : \rho > 0,$$

in the neighbourhood of $\rho = 0$.

Similar results can be obtained for the higher-order analogues to the DW test such as those studied by Wallis (1972) and Vinod (1973). For example, if the components of u in H_k are generated by the stationary, simple, fourth-order, autoregressive scheme,

$$u_t = \rho u_{t-4} + e_t, \quad |\rho| < 1,$$

where e is $E_o(n, I_n)$, then u is $E_o(n, \Sigma_4(\rho))$, where $\Sigma_4(\rho)$ has the form

$$\Sigma_4(\rho) = [(1-\rho)^2 I_n + \rho A_4 + \rho(1-\rho)C_4]^{-1}.$$

When n is an integer multiple of 4, A_4 and C_4 are defined as

$$A_4 = \bar{A}_1 \otimes I_4 \tag{5.4.10}$$

and

$$C_4 = \bar{C}_1 \otimes I_4, \tag{5.4.11}$$

where $\bar{A}_1$ and $\bar{C}_1$ are $({}^n/_4 \times {}^n/_4)$ matrices of the form of (4.2.29) and (4.2.30), respectively, and $\otimes$ denotes the Kronecker product.[15]

Approximating $\Sigma_4(\rho)$ by

$$\bar{\Sigma}_4(\rho) = [(1-\rho)^2 I_n + \rho A_4]^{-1}$$

allows us to apply Corollary 5.1.5. Hence, if the column space of X is spanned by some k characteristic vectors of A_4, the Wallis test statistic,

---

15. If n is not an integer multiple of 4, A_4 and C_4 are formed by first increasing n to the next integer multiple of 4, applying (5.4.10) and (5.4.11) and then reducing the resultant matrices to the required dimensions by removing an appropriate number of cross-diagonal elements.

$$d_4 = y'MA_4My/y'My,$$

with critical region $d_4 < c_0$, is an approximately UMPI test of $H_0: \rho = 0$ against $H_a: \rho > 0$, while, with an appropriate critical region of the form $d_4 < c_1$ or $d_4 > c_2$, it is an approximately UMPUI test of H_0 against the two-sided alternative $H_a': \rho \neq 0$.

The one-sided Wallis test can easily be shown to be locally UMPI in the neighbourhood of $\rho = 0$ under the assumption that e is $N(0, \sigma^2 I_n)$. By Corollary 5.1.2, this property also holds when e takes any other $E_0(n, I_n)$ distribution.

As an alternative to the DW test, Berenblut and Webb (1973) proposed the use of the statistic,

(5.4.12) $$g = y'(B - BX(X'BX)^{-1}X'B)y/y'My,$$

where B is the $n \times n$ matrix,

$$B = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & & 0 & 0 \\ 0 & -1 & 2 & & 0 & 0 \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ 0 & 0 & 0 & & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix},$$

in order to test for first-order autocorrelation in the disturbances of H_k. They considered both stationary autocorrelation as defined by (5.4.7) and autocorrelation generated according to the non-stationary scheme, (5.4.13) $$u_t = \rho u_{t-1} + e_t, \quad u_1 = e_1,$$

where ρ can take any value.

Throughout their analysis, Berenblut and Webb assumed normally distributed disturbances. When the column space of X is spanned by some k characteristic vectors of A_1 including the characteristic vector $\ell = (1,1,\ldots,1)'$, they found that their statistic, with critical region $g < c_o$, is an approximately UMP test of $H_0: \rho = 0$, against $H_a: \rho > 0$, in the stationary autoregressive scheme. The same test was shown to be locally UMPI as $\rho \to 1$, provided a constant dummy variable is included as a regressor. Against $H_a: \rho > 0$ in the non-stationary scheme, they demonstrated that their test is approximately UMP when the column space of X is spanned by some k characteristic vectors of B. They also showed it to be locally UMPI in the region of $\rho = 1$ against the same alternative hypothesis, but without any restriction on the design matrix.

If under H_k, the components of u are generated by the non-stationary, first-order, autoregressive scheme (5.4.13), where e is $E_o(n, I_n)$, then u is distributed $E_o(n, \Sigma_*(\rho))$, where $\Sigma_*(\rho)$ is of the form

(5.4.14) $$\Sigma_*(\rho) = [(1-\rho)^2 I_n + \rho B + \rho(1-\rho) C_*]^{-1}$$

and C_* is the $n \times n$ matrix C_1 with the unit value in the first row replaced by a zero. If (5.4.14) is approximated by

$$\bar{\Sigma}_*(\rho) = [(1-\rho)^2 I_n + \rho B]^{-1},$$

Corollaries 5.1.1, 5.1.2 and 5.1.5 allow Berenblut and Webb's results to be extended to:

<u>Corollary 5.1.8</u>. Suppose the linear regression model H_k holds with $u \sim \mathcal{E}_o(n,\Sigma(\rho))$, where $\Sigma(\rho)$ is of the form (5.4.8). If the column space of X is spanned by some k characteristic vectors of A_1 including ℓ, the Berenblut-Webb statistic, (5.4.12), with critical region $g < c_0$, is an approximately UMPI test of $H_0:\rho = 0$ against $H_a:\rho > 0$, while with an appropriate critical region of the form $g < c_1$ or $g > c_2$, it is an approximately UMPUI test of H_0 against the two-sided alternative $H_a':\rho \neq 0$. If $\mu\ell$, where μ is a non-zero scalar, is a column of X, the one-sided test is a locally UMPI test of $H_0:\rho = 0$ against $H_a:\rho > 0$ as $\rho \to 1$. For any given value of ρ, the distribution of g is the same for each $\mathcal{E}_o(n,\Sigma(\rho))$ distribution taken by u including the $N(0,\sigma^2\Sigma(\rho))$ distribution.

<u>Corollary 5.1.9</u>. Suppose the linear regression model H_k holds with $u \sim \mathcal{E}_o(n,\Sigma_*(\rho))$, where $\Sigma_*(\rho)$ is of the form (5.4.14). If the column space of X is spanned by some k characteristic vectors of B, the Berenblut-Webb statistic, (5.4.12), with critical region $g < c_0$, is an approximately UMPI test of $H_0:\rho = 0$ against $H_a:\rho > 0$, while with an appropriate critical region of the form $g < c_1$ or $g > c_2$, it is an approximately UMPUI test of H_0 against $H_a':\rho \neq 0$. The one-sided test is a locally UMPI test of $H_0:\rho = 0$ against $H_a:\rho > 0$ in the neighbourhood of $\rho = 1$. For any given value of ρ, the distribution of g is the same for each $\mathcal{E}_o(n,\Sigma_*(\rho))$ distribution taken by u including the $N(0,\sigma^2\Sigma_*(\rho))$ distribution.

With respect to the linear model H_k, where u is $N(0,\Sigma)$, Kadiyala (1970) examined the problem of testing $H_0:\Sigma = \sigma^2 I_n$ against the alternative $H_a:\Sigma = \sigma^2\Sigma_1$, where Σ_1 is a known $n \times n$ matrix. He noted that the disturbance vector, u, is not directly observable and proposed that his analysis should start with an observable random vector. His choice of the OLS residual vector enabled him to construct a test that is most powerful with respect to his transformed problem.

Fraser, Guttman and Styan (1976) argued that only the normed residuals, (5.3.1), are observable for tests of the values taken by Σ under H_k when u is distributed with mean 0 and covariance matrix Σ. Assuming $N(0,\Sigma)$ disturbances, they derived a most powerful test of $H_0:\Sigma = \sigma^2 I_n$ against $H_a:\Sigma = \sigma^2\Sigma_1$, where Σ_1 is fixed; this test being equivalent to Kadiyala's test. With respect to the problem of testing $H_0:\rho = 0$ against $H_a:\rho > 0$ in the stationary, first-order, autoregressive, disturbance scheme (5.4.7), a test that is locally most powerful in the neighbourhood of $\rho = 0$ was also constructed.

Corollary 5.1.2 implies that if the assumption that $u \sim N(0,\Sigma)$ is replaced by the wider assumption that $u \sim E_0(n,\Sigma)$, Kadiyala's test is most powerful within the class of tests invariant to the scale of the disturbances. It is also most powerful in the sense of Fraser et al. within this same class of tests. Further, their latter test is locally most powerful in the neighbourhood of $\rho = 0$ within this class of tests.

5. THE DISTRIBUTION OF STATISTICS NOT INVARIANT TO THE SCALE OF THE DISTURBANCES

In this section, we investigate the distributions of statistics which are not necessarily invariant to the scale of the disturbances in the general non-linear regression model, (5.2.1). Our findings allow us to discuss the validity of tests not invariant to the scale of the disturbances under H_k and H_k^{-1}, and to consider the power properties of tests not studied in the previous section.

Let $\mathcal{E}_1(n,\Sigma)$ denote an $\mathcal{E}(n,\Sigma)$ distribution with a joint density function, $g(z)$, of the form,

$$(5.5.1) \qquad g(z) = \int_o^\infty (2\pi\tau^2)^{-n/2}|\Sigma|^{-\frac{1}{2}}\exp\{-z'\Sigma^{-1}z/2\tau^2\}dF(\tau),$$

where $F(\tau)$ is a distribution function supported on (o,∞).[16] In this and the subsequent section, we shall often confine our attention to v taking an $\mathcal{E}_1(n,\Sigma)$ distribution. The transition from assuming $\mathcal{E}_o(n,\Sigma)$ disturbances to assuming $\mathcal{E}_1(n,\Sigma)$ disturbances is not particularly restrictive, especially in time series problems, since property IX implies that a sufficient condition for v to be $\mathcal{E}_1(n,\Sigma)$ is that it be an n observation sample from a spherically invariant stochastic process.

If $w = (w_1,\ldots,w_n)'$ and $x = (x_1,\ldots,x_n)'$ are n-dimensional vectors, let

$$w < x$$

denote

$$w_1 < x_1,\ w_2 < x_2,\ \ldots,\ w_n < x_n.$$

<u>Theorem 5.2</u>. Suppose that with respect to the non-linear regression model (5.2.1), $s(y)$ is a j-dimensional, Borel-measurable function of y with a joint distribution function,

---

16. Note that if $z \sim \mathcal{E}_1(n,\Sigma)$, $\Pr(z=0) = 0$ follows from $F(\tau)$ being supported on (o,∞) and hence $z \sim \mathcal{E}_o(n,\Sigma)$.

$$F_s(s;f,\theta,\tau^2,\Gamma) = \Pr(s(y)<s),$$

when v is distributed $N(0,\tau^2\Gamma)$ and where s is a j-dimensional vector and Γ is an $m \times m$ positive definite matrix. If v follows any other $E_1(m,\Gamma)$ distribution, the joint distribution function of $s(y)$ has the form

$$F_s^*(s;f,\theta,\Gamma) = \int_o^\infty F_s(s;f,\theta,\tau^2,\Gamma)\,dF(\tau),$$

where $F(\cdot)$ is a distribution function supported on (o,∞).

Proof: Define the j-dimensional, Borel-measurable function of v, $r(v)$, by

$$r(v) = s(f(\theta,v)).$$

When v is distributed $N(0,\tau^2\Gamma)$, the joint distribution function of $s(y)$ is

$$\begin{aligned}
&F_s(s;f,\theta,\tau^2,\Gamma) \\
&= \underset{r(v)<s}{\int\cdots\int} (2\pi\tau^2)^{-m/2}|\Gamma|^{-\frac{1}{2}}\exp\{-v'\Gamma^{-1}v/2\tau^2\}dv_1 \ldots dv_m,
\end{aligned}$$

while when v takes any other $E_1(m,\Gamma)$ distribution, the joint distribution function of $s(y)$ is

$$\begin{aligned}
&F_s^*(s;f,\theta,\Gamma) \\
&= \underset{r(v)<s}{\int\cdots\int} \int_o^\infty (2\pi\tau^2)^{-m/2}|\Gamma|^{-\frac{1}{2}}\exp\{-v'\Gamma^{-1}v/2\tau^2\}dF(\tau)\,dv_1 \ldots dv_m \\
&= \int_o^\infty \underset{r(v)<s}{\int\cdots\int} (2\pi\tau^2)^{-m/2}|\Gamma|^{-\frac{1}{2}}\exp\{-v'\Gamma^{-1}v/2\tau^2\}dv_1 \ldots dv_m\,dF(\tau) \\
&= \int_o^\infty F_s(s;f,\theta,\tau^2,\Gamma)\,dF(\tau);
\end{aligned}$$

the second equality following from Fubini's Theorem.

Like Theorem 5.1, Theorem 5.2 is versatile. Again we shall restrict our attention to its application to two special cases of (5.2.1), namely H_k and H_k^{-1}.

Theorem 5.2 implies that with respect to either of the linear regression models, H_k or H_k^{-1}, when u takes an $E_1(n,\Sigma)$ distribution, the size of any test with a Borel-measurable test statistic, is a weighted average of the sizes of the test for different positive values of σ^2 when u is $N(0,\sigma^2\Sigma)$ under the null hypothesis. If under normality, the test is an exact test in the sense that its actual size is always equal to the nominal significance level for all non-zero values of σ^2, Theorem 5.2 implies it remains an exact test when u takes any other $E_1(n,\Sigma)$ distribution. Note that such a case does not necessarily imply that the test statistic, itself, is invariant to non-zero values of σ^2. If the test statistic is invariant to $\sigma^2 > 0$,[17] Theorem 5.1 implies that the test is exact for all $E_o(n,\Sigma)$ distributions followed by u.

A more interesting case is that of a non-exact test which is considered to have a size sufficiently close to the nominal significance level for all positive values of σ^2 when u is

---

17. The author has been unable to find an example of a test proposed in the literature that is based on a test statistic which is not invariant to non-zero values of σ^2 but whose distribution function under the null hypothesis is. It is easy to construct trivial examples of such tests. For example, with respect to H_k define the test statistic d_1^* as follows:

$$d_1^* = \begin{cases} d_1 & \text{if } d_1 \neq 1 \\ y'My & \text{if } d_1 = 1, \end{cases}$$

where d_1 is the DW statistic, (5.4.9). Clearly, d_1^* has the same distribution as d_1 under $H_0{:}H_k$ holds with $u \sim N(0,\sigma^2 I_n)$, but d_1^*, itself, is not invariant to positive values taken by σ^2.

$N(0,\sigma^2\Sigma)$, for it to be a worthwhile test. From Theorem 5.2, it follows that when, under the null hypothesis, u takes any other $\mathcal{E}_1(n,\Sigma)$ distribution, such a test will also have a size sufficiently near the nominal significance level for the test to be of practical value.

A number of asymptotic tests for serial correlation in the disturbances of H_k^{-1} fall into this latter category of tests. Examples are the h-test and t-test suggested by Durbin (1970a), the procedure proposed by Box and Pierce (1970) for testing the disturbances of autoregressive schemes without exogenous regressors, the tests for autoregressive and moving average disturbances developed by Fitts (1973) and the various tests that have been constructed, using Silvey's (1959) Lagrange multiplier approach, by Breusch (1978) and Godfrey (1978a, 1978c, 1978d). Godfrey's (1973, 1974) tests for model misspecification under H_k and H_k^{-1} provide further examples of this latter type of test.

In general, such tests are asymptotic tests whose test statistics, under H_0 and the usual normality assumption, have small sample distributions which are dependent on the scale of the disturbances, but have asymptotic distributions which are independent of the disturbances' scale. The following corollary to Theorem 5.2 demonstrates that when the disturbances, $u(n)$, $n = 1, 2, \ldots,$ take any consistent sequence of $\mathcal{E}_1(n,\Gamma(n))$ distributions, the asymptotic distribution of such a test statistic is the same as that when the disturbances take the corresponding series of $N(0,\sigma^2\Gamma(n))$ distributions. The word "consistent" is used here to mean that for every $n_1 > n_2$, the distribution of $u(n_2)$ is the marginal distribution of the first n_2 components of $u(n_1)$. An immediate implication of this result is that the asymptotic tests listed above can be viewed as valid large sample tests when the disturbances take an appropriate $E_1(n,\Sigma)$ distribution.

Consider the consistent series of non-linear regression models,

$$(5.5.2) \qquad y(n) = f_n(\theta, v(m+n)), \qquad n=k,\ k+1,\ \ldots,$$

where for each value of n, $y(n)$ is an observable, n-dimensional random vector, f_n is a known, Borel measurable, vector function such that

$$f_n : R^k \times R^{n+m} \to R^n,$$

θ is a k-dimensional vector of parameters, $v(n+m)$ is an unobservable (n+m)-dimensional random vector and m is a constant integer such that $m > -k$.

Corollary 5.2.1. Suppose that with respect to the consistent series of non-linear regression models, (5.5.2),

$$s_n(y(n)), \qquad n=k,\ k+1,\ \ldots,$$

is a sequence of j-dimensional, Borel measurable functions of $y(n)$ each with joint distribution function,

$$F^n_{s_n}(s_n; f_n, \theta, \tau^2, \Gamma(m+n)) = Pr(s_n(y(n)) < s_n),$$

when $v(n)$, $n = 1, 2, \ldots$, take the consistent sequence of $N(0,\tau^2\Gamma(n))$ distributions. If $\{s_n(y(n))\}_{n=k}^{\infty}$ converges in distribution to the j-dimensional random vector s with distribution function,

$$F_s(s; \{f_n\}, \theta, \{\Gamma(n)\}),$$

that is invariant to the positive values taken by τ^2, then it also converges in distribution to s when $v(n)$, $n = 1, 2, \ldots$, take the corresponding sequence of $E_1(n,\Gamma(n))$ distributions.

<u>Proof</u>: Note that because $v(n)$, $n = 1, 2, \ldots,$ take the consistent sequence of $E_1(n,\Gamma(n))$ distributions, the joint density function of $v(n)$ is of the form (5.5.1) with a common distribution function $F(\cdot)$ supported on (o,∞).

Let

$$\bar{F}^n_{s_n}(s_n;f_n,\theta,\Gamma(m+n)) = \Pr(s_n(y(n))<s_n)$$

be the joint distribution function of $s_n(y(n))$ when $v(m+n)$ is distributed $E_1(m+n,\Gamma(m+n))$. For each value of s that is a point of continuity of $F_s(s;\{f_n\},\theta,\{\Gamma(n)\})$ we have,

$$\begin{aligned}
&\lim_{n\to\infty} \bar{F}^n_{s_n}(\, s;\, f_n,\theta,\Gamma(m+n))\\
&= \lim_{n\to\infty} \int_o^\infty F^n_{s_n}(\, s;\, f_n,\theta,\tau^2,\Gamma(m+n))dF(\tau)\\
&= \int_o^\infty \lim_{n\to\infty} F^n_{s_n}(\, s;\, f_n,\theta,\tau^2,\Gamma(m+n))dF(\tau)\\
&= \int_o^\infty F_s(s;\{f_n\},\theta,\{\Gamma(n)\})dF(\tau)\\
&= F_s(s;\{f_n\},\theta,\{\Gamma(n)\})
\end{aligned}$$

as required. The first equality is an application of Theorem 5.2 while the second follows from Lebesgue's bounded convergence theorem.

Occasionally, tests designed for use on H_k with normally distributed disturbances, such as those listed in (i)-(ix) of Section 3, are applied to the lagged dependent regressor model, H_k^{-1}. For specific tests, a good example being the DW test, there is a growing debate in the literature as to whether such procedures are appropriate.[18] Theorem 5.2 implies that the size of such a test when $\tilde{u}$ of H_k^{-1} takes an $E_1(n+p,\Sigma_o)$ distribution is a weighted average of the test's sizes for different positive values of σ^2 when $\tilde{u} \sim N(0,\sigma^2\Sigma_o)$.

Theorem 5.2 also implies that the power of such tests for serial correlation or heteroscedasticity under an alternative hypothesis of $\tilde{u} \sim E_1(n+p,\Sigma_1)$, is a weighted average of the corresponding powers for different positive values of σ^2 when $\tilde{u} \sim N(0,\sigma^2\Sigma_1)$. An identical conclusion holds for the asymptotic tests for serial correlation in the disturbances of H_k^{-1}, listed above. The power properties of a number of these tests, assuming normally distributed disturbances, have been studied by Taylor and Wilson (1964), Park (1972, 1975), Maddala and Rao (1973), Kenkel (1974, 1975), Spencer (1975) and Godfrey and Tremayne (1978). In view of the above comments, one might expect similar powers, and hence similar conclusions, regarding the ranking of tests' performances, if in these studies, non-normal, elliptically symmetric disturbances had been used.

We now turn our attention to the linear regression model H_k. In the previous section, we discussed the power properties of tests for serial correlation and heteroscedasicity in the disturbances of H_k. With the aid of Theorem 5.2, we shall consider the power properties of the various tests for misspecification and outlying observations listed in Section 3.

---

18. Durbin and Watson (1950), Nerlove and Wallis (1966) and Park (1976) have argued that the DW test should not be used in this situation, while Taylor and Wilson (1964) and Kenkel (1974, 1976) have recommended the opposite.

In general, these are tests of the null hypothesis,

(5.5.3) $$H_0: y \sim N(X\beta, \sigma^2\Sigma),$$

against the alternative,

(5.5.4) $H_a: y$ is generated by a non-linear model of the form (5.2.1) with $v \sim N(0, \tau^2\Gamma)$,

where Γ is an $n \times n$, nonstochastic, positive definite matrix. Tests for outliers are possible exceptions; occasionally the alternative hypothesis does not require the outlying observation to be normally distributed. Our concern is with the effect on the power of tests of the usual normality assumption being widened to an appropriate elliptical symmetry assumption. Consequently, we shall not discuss test statistics, which under the usual alternative hypothesis, are functions of non-normal disturbances.

Consider a test statistic, $s(y)$, which, together with an appropriate critical region, is used to test (5.5.3) against (5.5.4). If under H_0, the distribution of $s(y)$ is invariant to $\sigma^2 > 0$ then, as noted previously in this section, $s(y)$ has the same null distribution for any $E_1(n,\Sigma)$ distribution taken by $y - X\beta$. If $s(y)$, itself, is invariant to positive values of σ^2, the null distribution of $s(y)$ is also invariant to the type of $E_0(n,\Sigma)$ distribution followed by $y - X\beta$. If under H_a, the distribution of $s(y)$ is invariant to positive values taken by τ^2, Theorem 5.2 implies $s(y)$ has the same alternative distribution and hence the same power, for any $E_1(n,\Gamma)$ distribution taken by v. This property also holds for any $E_0(n,\Gamma)$ distribution, if $s(y)$, itself, is invariant to $\tau^2 > 0$.

On the other hand, if under H_a the distribution of $s(y)$ is not invariant to $\tau^2 > 0$, then Theorem 5.2 implies that under the alternative hypothesis,

$$H_a^{\prime}: y \text{ is generated by a non-linear model of the form (5.2.1) with } v \sim \mathcal{E}_1(n,\Gamma),$$

the distribution of $s(y)$ can be viewed as a mixture of the distributions taken by $s(y)$ under H_a for different positive values of τ^2. Hence under $H_a^{\prime}$, the test's power is a weighted average of its power for different values of $\tau^2 \in (o,\infty)$ under H_a; the weights depending upon the particular type of $\mathcal{E}_1(n,\Gamma)$ distribution v follows.

Obviously, if a particular test of H_0 against H_a listed in Section 3 has an optimal power property for all values of $\tau^2 \in (o,\infty)$, then this same optimal property will hold when the test is used to test the null hypothesis,

$$H_0^{\prime}: y - X\beta \sim \mathcal{E}_0(n,\Sigma),$$ [19]

against $H_a^{\prime}$. For example, Lehmann (1959, p.265) showed that under the maintained hypothesis of H_k with X of full column rank and $u \sim N(0,\sigma^2 I_n)$, the familiar F-test used to test the validity of linear restrictions of the form

$$A\beta = 0,$$

where A is an $r \times k$ nonstochastic matrix of rank r, is UMPI.

---

19. Recall that we are assuming the test statistic is invariant to the scale of the disturbances under H_0 and hence the test has the same size for any $\mathcal{E}_0(n,\Sigma)$ distribution followed by u.

This property will also hold when u takes any other $E_1(n, I_n)$ distribution.

Finally, we close this section with some comments on the distributions of the coefficient of determination, R^2, and the coefficient of determination adjusted for degrees of freedom, $\bar{R}^2$, with respect to H_k, when the disturbances are spherically symmetric. For $N(0, \sigma^2 I_n)$ disturbances and selected values of X, β and σ^2, Koerts and Abrahamse (1969, Chapter 8) calculated values of the distribution function of R^2. In a similar manner, Ebbeler (1975) calculated probabilities of making a correct selection when the true regression model is one of two alternative specifications, and the model with the larger $\bar{R}^2$ value is chosen. Theorem 5.2 implies that the values of such probabilities, when u takes any non-normal $E_1(n, I_n)$ distribution, are weighted averages of the corresponding probabilities for different values of σ^2 in the range $(0, \infty)$, when u is $N(0, \sigma^2 I_n)$.

6. THE DISTRIBUTION OF REGRESSION ESTIMATORS

In this section we examine the consequences of Theorems 5.1 and 5.2 for the distributions of a number of better known estimators of β in the regression models H_k and H_k^{-1} when the disturbances are elliptically symmetric. In addition, we consider the distributions of estimators of nuisance parameters connected with the disturbance term of H_k. Special attention is given to various estimators of the first-order autocorrelation coefficient, ρ, in the case when the regression disturbances follow the stationary, first-order, autoregressive scheme (5.4.7).

The following corollary, which is an immediate consequence of Corollary 5.2.1, allows us to say something about the distribution of almost any estimator of β in H_k when the disturbances are $\mathcal{E}_1(n,\Sigma)$.

Corollary 5.2.2. With respect to the linear regression model H_k, with $u \sim \mathcal{E}_1(n,\Sigma)$, the joint distribution function of any Borel-measurable estimator of β is the weighted average of the estimator's joint distribution functions for different positive values of σ^2 when the disturbances are $N(0,\sigma^2\Sigma)$.

A parallel result holds for the lagged dependent variable regression model H_k^{-1}.

Corollary 5.2.3. With respect to the regression model H_k^{-1}, with $\tilde{u} \sim \mathcal{E}_1(m,\Gamma)$, the joint distribution function of any Borel-measurable estimator of $\beta^* = (\alpha,\beta_1,\ldots,\beta_k)'$, is a weighted average of the estimator's joint distributions for different positive values of σ^2 when $\tilde{u} \sim N(0,\sigma^2\Gamma)$, where $m = n + p$ and Γ is $m \times m$.

Hence, when the disturbance vector of H_k or H_k^{-1} takes an $\mathcal{E}_1(n,\Sigma)$ distribution, any linear unbiased, any linear biased or any well-behaved non-linear estimator will have very similar properties to those of the same estimator when the disturbance term is normally distributed. Unfortunately Corollaries 5.2.2 and 5.2.3 don't tell us how similar. The following result, which is independent of Theorem 5.2 and its corollaries, is of some assistance on this latter point.

Theorem 5.3. Suppose $\tilde{\beta}(y)$ is a Borel-measurable estimator of β in the linear regression model H_k.

(i) If, when $u \sim N(0,\sigma^2\Sigma)$, $\tilde{\beta}(y)$ is unbiased for all positive values of σ^2, then, for any other $E_1(n,\Sigma)$ distribution the disturbances may follow, $\tilde{\beta}(y)$ is unbiased provided its first order moments exist.

(ii) If, when $u \sim N(0,\sigma^2\Sigma)$, $\tilde{\beta}(y)$ is unbiased and has a covariance matrix of the form

$$\gamma(\sigma^2)C$$

for all positive values of σ^2, where $\gamma(\cdot)$ is a positive scalar function on (o,∞) and C is a $k \times k$ positive semi-definite matrix independent of σ^2, then for any other $E_1(n,\Sigma)$ distribution the disturbances may follow, the covariance matrix of $\tilde{\beta}(y)$ is of the form $\delta^2 C$ where $\delta^2 > 0$, provided the second order moments of $\tilde{\beta}(y)$ exist.

<u>Proof</u>: (i) For $u \sim E_1(n,\Sigma)$ and assuming $E(\tilde{\beta}(y))$ exists,

$$\begin{aligned}
&E(\tilde{\beta}(y)) \\
&= \int_{R^n} \tilde{\beta}(X\beta+u) \int_o^\infty (2\pi\tau^2)^{-n/2} |\Sigma|^{-\frac{1}{2}} \exp\{-u'\Sigma^{-1}u/2\tau^2\} dF(\tau) du \\
&= \int_o^\infty \int_{R^n} \tilde{\beta}(X\beta+u) (2\pi\tau^2)^{-n/2} |\Sigma|^{-\frac{1}{2}} \exp\{-u'\Sigma^{-1}u/2\tau^2\} du dF(\tau) \\
&= \beta;
\end{aligned}$$

the second equality following from Fubini's Theorem while the third is a consequence of $\tilde{\beta}(y)$ being unbiased for normal disturbances.

(ii) Let $B(X,\beta,u)$ denote the $k \times k$ matrix,

$$B(X,\beta,u) = (\tilde{\beta}(X\beta+u)-\beta)(\tilde{\beta}(X\beta+u)-\beta)'.$$

For $u \sim E_1(n,\Sigma)$ and assuming $\tilde{\beta}(y)$ is unbiased and that its second order moments exist, then

$$
\begin{aligned}
&\mathrm{Var}(\tilde{\beta}(y)) \\
&= \int_{R^n} B(X,\beta,u)\int_o^\infty (2\pi\tau^2)^{-n/2}|\Sigma|^{-\frac{1}{2}}\exp\{-u'\Sigma^{-1}u/2\tau^2\}dF(\tau)du \\
&= \int_o^\infty\int_{R^n} B(X,\beta,u)\,(2\pi\tau^2)^{-n/2}|\Sigma|^{-\frac{1}{2}}\exp\{-u'\Sigma^{-1}u/2\tau^2\}dudF(\tau) \\
&= \int_o^\infty \gamma(\tau^2)\,CdF(\tau) \\
&= \delta^2 C,
\end{aligned}
$$

where δ^2 is a positive scalar since $\gamma(\tau^2)$ is a positive function on (o,∞).

Again, a parallel result also holds for the lagged dependent variable model.

As well as applying to the usual OLS, GLS, restricted least squares and instrumental variable estimators, the results established in this section also apply to an important class of estimators known as pretest or sequential estimators. Such estimators occur when the same data are used to select a particular final specification of the model by prior testing as well as to estimate the selected model's parameters.

The distributions of these estimators under the usual assumption of normally distributed disturbances have been the subject of a number of papers in recent years.[20] Conclusions that are independent of the scale of the disturbances and based on pretest estimator distributions calculated in these studies, are therefore equally valid when the disturbance vector takes an appropriate $E_1(n,\Sigma)$ distribution.

---

20. For example, see the references in Wallace (1977).

Obviously, results analogous to Corollaries 5.2.2 and 5.2.3 can be obtained for estimators of nuisance parameters connected with the characteristic matrix Σ, for $E_1(n,\Sigma)$ distributed regression disturbances. An example of such a nuisance parameter is the first-order autocorrelation coefficient, ρ, when the disturbances follow the stationary, first-order, autoregressive scheme, (5.4.7). On the other hand, Theorem 5.1 implies that when a given estimator of such a nuisance parameter is invariant to the positive values taken by σ^2 under H_k with $N(0,\sigma^2\Sigma)$ disturbances, the estimator's distribution remains unchanged if u takes any other $E_0(n,\Sigma)$ distribution. The remainder of this section is devoted to identifying some of the better known examples of such special cases.

Recently, Magnus (1978) derived the first-order conditions for the maximum likelihood estimators of the linear regression model H_k,[21] where $u \sim N(0,\Omega)$ and $\Omega = \sigma^2\Sigma$ is a function of unknown parameters. When the elements of Ω are twice differentiable functions of a finite and constant number of parameters, $\theta_1, \ldots, \theta_m$, these first-order maximum likelihood conditions are

$$\hat{\beta} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y, \tag{5.6.1}$$

$$\mathrm{tr}\left(\frac{\partial\Omega^{-1}}{\partial\theta_j}\Omega\right)\Big|_{\theta=\hat{\theta}} = e'\left(\frac{\partial\Omega^{-1}}{\partial\theta_j}\right)\Big|_{\theta=\hat{\theta}} e, \qquad j=1,\ldots,m, \tag{5.6.2}$$

where $e = y - X\hat{\beta}$ and $\hat{}$ denotes the maximum likelihood estimator.

If $\theta_1 = \sigma^2$ and $\Omega = \sigma^2\Sigma(\theta_2,\ldots,\theta_m)$, (5.6.1) becomes,

$$\hat{\beta} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y,$$

---

21. For the remainder of this section, the design matrix of H_k will be assumed to be of full column rank.

the first equation of (5.6.2) becomes

$$\hat{\sigma}^2 = e'\hat{\Sigma}^{-1}e/n,$$

and the last $m - 1$ equations of (5.6.2) become

$$\text{(5.6.3)} \qquad \mathrm{tr}\left(\frac{\partial \Sigma^{-1}}{\partial \theta_j}\Sigma\right)\Big|_{\theta=\hat{\theta}} = e'\frac{\partial \Sigma^{-1}}{\partial \theta_j}\Big|_{\theta=\hat{\theta}} e/\hat{\sigma}^2, \qquad j=2,\ldots,m.$$

The maximum likelihood estimates of $\theta_2, \ldots, \theta_m$ are determined by the $m - 1$ equations (5.6.3). Since these equations are invariant to the positive values taken by σ^2, the maximum likelihood estimators of the unknown parameters of Σ also will be invariant to the positive values taken by σ^2, as required for the application of Theorem 5.1.

The consequences of this result are reasonably wide ranging. For example, when the components of u are generated by an ARMA (Autoregressive Moving Average) process whose order is known, the maximum likelihood estimators of the ARMA parameters have the same distribution whether the errors of the process are $N(0,\sigma^2 I_n)$ or follow any other $E_0(n,I_n)$ distribution.

When the components of u in H_k are generated by the stationary, first-order, autoregressive scheme, (5.4.7), with $e \sim N(0,\sigma^2 I_n)$, it is a simple task to prove that the iterative estimator of ρ proposed by Cochrane and Orcutt (1949) is invariant to the positive values of σ^2, provided the initial estimate of ρ is invariant to values of σ^2. Not so obvious is the fact that the estimator of ρ proposed by Durbin (1960) also is invariant to the positive values of σ^2.

Durbin's estimator of ρ in (5.4.7) is simply the OLS estimator of ρ in the equation

$$y = \rho y_{-1} + X\gamma + X_{-1}\delta + \varepsilon$$

(5.6.4) $$= \rho y_{-1} + X^*\eta + \varepsilon,$$

where $$y_{-1} = (y_0, \ldots, y_{n-1})',$$

γ and δ are respectively k- and j-dimensional vectors of unknown parameters, X_{-1} is the $n \times j$ matrix made up of columns,

$$x_{-1i} = (x_{0i}, \ldots, x_{n-1i})', \qquad i=1,\ldots,k,$$

but with the minimum number of columns deleted so as to make $[y_{-1} \vdots X \vdots X_{-1}]$ a full column rank matrix,

$$X^* = [X \vdots X_{-1}]$$

and

$$\eta = \begin{bmatrix} \gamma \\ \cdots \\ \delta \end{bmatrix}.$$

The OLS estimator of $\eta^* = (\rho \vdots \eta')'$ is

$$\hat{\eta}^* = \begin{bmatrix} y_{-1}'y_{-1} & y_{-1}'X^* \\ X^{*\prime}y_{-1} & X^{*\prime}X^* \end{bmatrix}^{-1} \begin{bmatrix} y_{-1}'y \\ X^{*\prime}y \end{bmatrix}.$$

Thus the OLS estimator of ρ in (5.6.4) is

$$\hat{\rho} = [(y_{-1}'M^*y_{-1})^{-1} \vdots -(y_{-1}'M^*y_{-1})^{-1}y_{-1}'X^*(X^{*\prime}X^*)^{-1}] \begin{bmatrix} y_{-1}'y \\ X^{*\prime}y \end{bmatrix}$$

(5.6.5) $$= y_{-1}'M^*y/y_{-1}'M^*y_{-1},$$

where $$M^* = I_n - X^*(X^{*\prime}X^*)^{-1}X^{*\prime}.$$

Since

$$M^*X^* = 0,$$

$$M^*x^*_{.i} = 0,$$

where $x^*_{.i}$ is any column of X or X_{-1}. Substitution of (5.2.3) into (5.6.5) yields

$$\hat{\rho} = u'_{-1}M^*u/u'_{-1}M^*u_{-1},$$

which is obviously invariant to the scale of u thus permitting the application of Theorem 5.1.

7. CONCLUSIONS

In this chapter, a number of results were derived concerning statistics of an n-dimensional random vector, y, which, itself, is a function of a disturbance vector u. The central theme is that the distribution of any function of y, whether it be an estimator of some unknown parameter, a test statistic or any other statistic, when u takes an $E_1(n,\Gamma)$ distribution can be viewed as a weighted average of the distributions taken by the function for different values of σ^2 when u is $N(0,\sigma^2\Gamma)$. In the special case when the function, itself, is invariant to the positive values taken by σ^2 assuming $u \sim N(0,\sigma^2\Gamma)$, the function has the same distribution for any $E_0(n,\Gamma)$ distribution u may follow.

The application of this remarkable property to the parameter estimators and statistical tests associated with the usual linear regression model, H_k, and the lagged dependent variable regression model, H_k^{-1}, allows us to conjecture that there would be few significant changes in the distributions of such estimators and the properties of such tests when the usual assumption of normally distributed disturbances is replaced by an appropriate elliptical symmetry assumption.

Since the key results of this chapter all relate to the general non-linear regression model (5.2.1), it is clear that this property also has similar applications to these models. In addition, it obviously has applications to simultaneous equation models and time series models whose random components are often assumed to be generated by normally distributed disturbances; a prominent example being the ARIMA (Autoregressive Integrated Moving Average) processes studied by Box and Jenkins (1970). These applications are not pursued here because they fall beyond the scope of this thesis.

CHAPTER 6

TESTING FOR AUTOCORRELATION USING LINEAR UNBIASED REGRESSION RESIDUALS WITH SCALAR COVARIANCE MATRICES[1]

1. INTRODUCTION

Much of the recent literature on the subject of testing for autocorrelation in linear regression disturbances has been concerned with the problem of the inconclusive region in the DW bounds test. The inconclusive region results from the fact that the joint probability distribution of the OLS residuals, in the usual linear regression model, is dependent on the design matrix. In order to overcome this difficulty, Theil (1965, 1968, 1971) proposed the use of residuals whose joint distribution is independent of the design matrix, in tests for serial correlation. He derived his BLUS (Best Linear Unbiased with Scalar covariance matrix) residual vectors and suggested their use in the von Neumann ratio as an alternative to the DW bounds test.

A number of authors followed Theil's lead and also proposed tests based on residuals distributed independently of the regressors. Durbin (1970b) constructed an exact test based on the DW statistic calculated using residuals whose joint distribution is independent of the design matrix. Sims (1975) later suggested a simple modification to this test in order to improve its power properties.

---

1. Throughout this chapter, residual vectors are treated as estimators of the particular value taken by the disturbance vector. The terms "residuals" and "disturbance estimator" are used interchangably in this context.

Golub and Styan (1973) showed that Golub's (1965) LUSH residual vector has a scalar covariance matrix and hence can be used in the von Neumann ratio. Phillips and Harvey (1974) proposed a simple test for serial correlation based on recursive residuals which also are LUS (Linear Unbiased with Scalar covariance matrix). More recently, Dent and Styan (1978) introduced a test based on Tiao and Guttman's (1967) BAUS (Best Augmented Unbiased with Scalar covariance matrix) residuals.

The relative power properties of many of the above tests have been investigated by various authors. Koerts and Abrahamse (1969) found that for selected design matrices, Theil's BLUS test has lower power than the DW test using the exact critical value. This conclusion has been verified by a number of studies including those of Dubbelman (1972) and L'Esperance and Taylor (1975).[2] Ward (1973) found that the power of tests based on BLUS residuals is frequently lower than the power of tests based on LUSH residuals, while Phillips and Harvey (1974) found little to choose between tests based on recursive residuals and BLUS residuals in terms of power alone. Dent and Styan (1978) reported a similar conclusion regarding the relative power properties of tests based on BLUS and BAUS residuals.

Theil's BLUS residuals are "best" in the sense that they minimize the expected sum of squares of the estimation errors for selected $n - k$ disturbances, but from the point of view of testing for auto-

---

2. Dubbelman, Abrahamse and Louter (1978) have pointed out that L'Esperance and Taylor's algorithms for calculating the power functions of the AK test and Durbin's test appear to be incorrect. However, their algorithms for the DW and BLUS tests, as presented, are correct and their results for these tests support Koerts and Abrahamse's findings.

correlation in these disturbances, their performance is disappointing. Abrahamse and Koerts (1969, 1971) hypothesised that this is due to the scalar covariance matrix restriction. They constructed a disturbance estimator (hereafter called the AK (Abrahamse-Koerts) residual vector) with a non-scalar covariance matrix that is chosen *a priori* and proposed its use in the DW statistic as a test for autocorrelation in the disturbances.

Empirical power studies by Abrahamse and Louter (1971), Dubbelman (1972), Dubbelman, Abrahamse and Louter (1978) and Dent and Cassing (1978) suggest that, for a number of design matrices that might be encountered in economic time series analysis, tests based on the AK residual vector are superior to those based on BLUS residuals and have similar power properties to the exact DW test. On the other hand, these studies have also identified design matrices for which the AK test has relatively poor power properties.

The aim of this chapter is to investigate the class of LUS disturbance estimators[3] with respect to the usual linear regression model,

$$y = X\beta + u, \tag{6.1.1}$$

where y is an observable, n-dimensional random vector, X is an $n \times k$ nonstochastic matrix, β is a k-dimensional vector of unknown

---

3. In order to facilitate a generalization to elliptically symmetric disturbances, we shall call a linear disturbance estimator, v, unbiased if, when the disturbance vector u is elliptically symmetric, v is also elliptically symmetric, although not necessarily with the same characteristic matrix. We shall call v a LUS residual vector, if when u is $E_o(n,I_n)$, v takes an $E_o(m,I_m)$ distribution.

parameters, and u is an n-dimensional disturbance vector. We shall attempt to choose the "best" LUS residual test for first-order autoregressive disturbances in typical economic time series linear regression models.

In Section 2, the literature concerning the problem of testing for a given disturbance covariance matrix is reviewed. Then, Corollary 5.1.3 is used to construct a UMPI test for a given characteristic matrix, when the disturbances are assumed to be elliptically symmetric. In Section 3, we find that against the alternative hypothesis of first-order autoregressive disturbances, the LUS test which "best"[4] approximates the UMPI test of Section 2, is identical to the AK test. Implications of this result are discussed and the extension to the AK test proposed by Dubbelman (1972) is analysised in Section 4. Conclusions are presented in the final section.

2. TESTING FOR A GIVEN CHARACTERISTIC MATRIX OF REGRESSION DISTURBANCES

With respect to the observable random vector y determined by the linear regression model, (6.1.1), with

$$u \sim \mathcal{E}_o(n,\Sigma), \tag{6.2.1}$$

consider the problem of testing the null hypothesis,

$$H_0: \Sigma = \Sigma_0,$$

---

4. "Best" for a particular class of design matrix.

against the alternative hypothesis,

$$H_a: \Sigma = \Sigma_1,$$

where Σ_0 and Σ_1 are given positive definite matrices.

A number of authors have studied this problem in the special case of normally distributed disturbances. Anderson (1948) analysed it for Σ_0 and Σ_1 of the form:

$$\Sigma_0 = \Psi^{-1},$$

$$\Sigma_1 = [\Psi+\lambda\Theta]^{-1}, \tag{6.2.2}$$

where Ψ is a given, $n \times n$, positive definite matrix, Θ is a given, $n \times n$ matrix and λ a scalar such that (6.2.2) is positive definite. He restricted his attention to tests that are similar. A test is similar if its critical region at the α level of significance,

$$\omega \subset R^n,$$

is such that, for all allowable distribution functions, F, of y under the null hypothesis,

$$\int_\omega dF(y) = \alpha,$$

as opposed to the usual assumption that

$$\int_\omega dF(y) \leqslant \alpha,$$

with equality for at least one distribution function.

For one-sided alternatives[5] and when the column space of the

---

5. I.e., when testing $\lambda = 0$ against $\lambda > 0$ or against $\lambda < 0$.

transformed design matrix $\Gamma'X$ is spanned by some k characteristic vectors of $\Gamma'\Theta\Gamma$, where Γ is an $n \times n$, nonsingular matrix such that

$$\Gamma'\Psi\Gamma = I_n$$

and

$$\Gamma\Gamma' = \Psi^{-1},$$

Anderson provides a test that is UMP within the class of similar tests. Clearly, this condition on the design matrix is likely to be satisfied only on rare occasions.

As noted in the previous chapter, Kadiyala (1970) constructed a test that, for normal disturbances, is MP with respect to the related, but different, problem in which the observed random vector is taken to be the OLS residual vector,

$$z = My,$$

instead of y, where

$$M = I_n - X(X'X)^{-1}X'.$$

Durbin and Watson (1971) considered the problem of testing the null hypothesis, $H_0: \rho = 0$, against the alternative, $H_a: \rho = \rho_1$, where ρ_1 is a given positive scalar, when the components of the disturbance vector are generated by the stationary, first-order, autoregressive scheme (5.4.7). They declared their preference for a theory of invariance approach to the problem, as opposed to restricting consideration to similar tests as favoured by Anderson (1948) and adopted by themselves in an earlier justification of the

DW test. After approximating the true characteristic matrix of the disturbances under H_a, (5.4.8), by

$$(6.2.3) \qquad \bar{\Sigma} = [(1-\rho)^2 I_n + \rho A_1]^{-1},$$

and assuming normal disturbances, they attempted to derive a MP invariant test, where invariance is with respect to transformations of the form

$$(6.2.4) \qquad y \to \gamma_o y + X\gamma,$$

where γ_o is a positive scalar and γ is a k-dimensional vector. Webb (1973) has pointed out a mistake in their proof and consequently the test they finally obtain is not MP invariant in all cases, the only exception being when the GLS residual vector, $\tilde{u}$, and z coincide.[6] Because Durbin and Watson's conclusions regarding optimal power properties of the DW test concern situations where $\tilde{u}$ and z coincide, these conclusions are unaffected.

As mentioned in Chapter 5, Fraser, Guttman and Styan (1976) showed that the normed residual vector, whose components are given by (5.3.1), is a directly observable random vector for the problem of testing the covariance matrix of the disturbance vector, u, in (6.1.1). They used this fact to obtain a MP test of H_0 against H_a when $\Sigma_0 = I_n$ and the regression disturbances are normally distributed; the test being equivalent to Kadiyala's test.

---

6. Using the notation of Durbin and Watson, the result of this error is that their rejection region $\{\hat{u}'\hat{u}/\hat{u}'B\hat{u}\}^{\frac{1}{2}(n-k)} > c$ should have been $\{z'z/\hat{u}'B\hat{u}\}^{\frac{1}{2}(n-k)} > c$, where $\hat{u}$ denotes the GLS residual vector. Therefore, the MPI test is $\hat{u}'B\hat{u}/z'z < c_1$ and not $\hat{u}'A\hat{u}/\hat{u}'\hat{u} < c_3$ as stated by Durbin and Watson.

There are also the contributions made by Kariya and Eaton (1977) and Kariya (1977) which have been thoroughly outlined in Section 4 of Chapter 5.

The aim of this section is to find a test of H_0 against H_a with optimal power properties. Of all the various approaches to the problem that are outlined above, we favour the use of the theory of invariance to obtain a UMPI test. Applying Neyman-Pearson methods to the class of tests similar under the null hypothesis, the technique used by Anderson, has been criticised by a number of prominent statisticians,[7] while Kadiyala's approach involves a plausible but non-rigorous change in the hypotheses being tested. We conjecture that a well-defined relationship exists between Fraser, Guttman and Styan's approach and that using the theory of invariance.[8]

We can assume without loss of generality that $\Sigma_0 = I_n$, since if $\Sigma_0 \neq I_n$, (6.1.1) can be transformed by premultiplying by the non-singular matrix $\Sigma_0^{-\frac{1}{2}}$ and the problem becomes one of testing $u \sim E_0(n, I_n)$ against $u \sim E_0(n, \Sigma_0^{-\frac{1}{2}}\Sigma_1(\Sigma_0^{-\frac{1}{2}})')$.

In the theorem which follows, we present a UMPI test of H_0 against H_a. In proving this theorem, we make use of Corollary 5.1.3 and the following lemma:

<u>Lemma 6.1.</u> If w is an $E_0(n,\Sigma)$ random vector, then

$$w/(w'w)^{\frac{1}{2}}$$

---

7. For example, see Lehmann and Stein (1948), Durbin and Watson (1971) and Fraser et al. (1976).

8. An investigation of the possible existence of such a relationship was considered to be beyond the scope of this thesis.

is also an $E_0(n,\Sigma)$ random vector.

Proof: $\Sigma^{-\frac{1}{2}}w$ is $E_0(n,I_n)$ and hence for any $n \times n$ orthogonal matrix, A,

$$\Sigma^{-\frac{1}{2}}w \quad \text{and} \quad A\Sigma^{-\frac{1}{2}}w$$

obey the same distribution laws. Since

$$\Sigma^{-\frac{1}{2}}w/(w'w)^{\frac{1}{2}} \quad \text{and} \quad A\Sigma^{-\frac{1}{2}}w/(w'w)^{\frac{1}{2}}$$

also obey the same distribution laws for any $n \times n$ orthogonal matrix A,

$$\Sigma^{-\frac{1}{2}}w/(w'w)^{\frac{1}{2}}$$

is $E_0(n,I_n)$ and the result follows.

Theorem 6.1. With respect to the linear regression model (6.1.1) and (6.2.1), let P be any $n \times n$ matrix such that

(6.2.5) $$PMP' = \begin{bmatrix} I_{n-k} & 0 \\ 0' & 0_k \end{bmatrix},$$

(6.2.6) $$PP' = P'P = I_n$$

and let P be partitioned as

(6.2.7) $$P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix},$$

where P_1 is $(n-k) \times n$ and P_2 is $k \times n$. For testing $H_0: \Sigma = I_n$ against $H_a: \Sigma = \Sigma_1$, where Σ_1 is a given, positive definite matrix, the test which rejects H_0 for small values of

(6.2.8) $$s = v'(P_1\Sigma_1 P_1')^{-1}v/v'v$$

is UMPI, where

$$v = P_1 z$$

and invariance is with respect to transformations of the form (6.2.4).

<u>Proof</u>: First we shall show that for any $n \times n$ matrix P such that (6.2.5) and (6.2.6) hold,

$$Pz/(z'z)^{\frac{1}{2}} \tag{6.2.9}$$

is a maximal invariant.

Obviously (6.2.9) is invariant to transformations of the form (6.2.4). Let $z_{(1)}$ and $z_{(2)}$ be OLS residual vectors from the (n×k) linear regression equations,

$$y_{(1)} = X\beta + u_{(1)},$$

and

$$y_{(2)} = X\beta + u_{(2)},$$

respectively. If

$$Pz_{(1)}/(z'_{(1)}z_{(1)})^{\frac{1}{2}} = Pz_{(2)}/(z'_{(2)}z_{(2)})^{\frac{1}{2}},$$

then

$$Mu_{(1)} = \delta Mu_{(2)}, \tag{6.2.10}$$

where δ is the positive scalar,

$$\delta = (z'_{(1)}z_{(1)})^{\frac{1}{2}}/(z'_{(2)}z_{(2)})^{\frac{1}{2}}.$$

(6.2.10) implies

$$u_{(1)} - \delta u_{(2)} = X\theta,$$

where

$$\theta = (X'X)^{-1}X'(u_{(1)}-\delta u_{(2)}).$$

Hence,

$$y_{(1)} - \delta y_{(2)} = X\{\theta+(1-\delta)\beta\}.$$

Thus $y_{(1)}$ and $y_{(2)}$ are related by a transformation of the form (6.2.4) and, therefore, (6.2.9) is a maximal invariant.

Note that

(6.2.11)
$$Pz = \begin{bmatrix} P_1 z \\ 0 \end{bmatrix},$$

since (6.2.5) and (6.2.6) imply

$$PM = \begin{bmatrix} P_1 \\ 0 \end{bmatrix}$$

and

(6.2.12)
$$P_1 u = P_1 z$$

follows from $P_1 M = P_1$. From (6.2.6) we have $P_1'P_1 = I_n - P_2'P_2$. Postmultiplying by M yields

(6.2.13)
$$P_1'P_1 = M .$$

Then (6.2.11), (6.2.12) and (6.2.13) imply

(6.2.14)
$$P_1 z/(z'z)^{\frac{1}{2}} = v/(v'v)^{\frac{1}{2}}$$

is a maximal invariant, where

$$v = P_1 z = P_1 u.$$

That $v/(v'v)^{\frac{1}{2}}$ takes an $E_0(n-k, I_{n-k})$ distribution under H_0 and an $E_0(n-k, P_1\Sigma_1P_1')$ distribution under H_a, follows from property III of Chapter 3 and Lemma 6.1.

Since (6.2.14) is a maximal invariant, all invariant test statistics can be expressed as functions of it, and Corollary 5.1.3 implies that the test which rejects H_0 for small values of

$$s = v'(P_1\Sigma_1P_1')^{-1}v/v'v$$

is UMPI.[9]

The UMPI test of Theorem 6.1 can be expressed in terms of OLS residuals and GLS residuals as the corollary below demonstrates. In proving this corollary, we use the following lemma from Webb (1973).[10]

Lemma 6.2.

$$V^{-1} - V^{-1}U(U'V^{-1}U)^{-1}U'V^{-1} = T(T'VT)^{-1}T',$$

where V is any $n \times n$ positive definite matrix and U and T are $n \times k$ and $n \times (n-k)$ matrices, respectively, such that if $W = (U \vdots T)$ then

$$W'W = WW' = I_n.$$

Corollary 6.1.1. Consider the problem of testing $H_0: \Sigma = I_n$ against $H_a: \Sigma = \Sigma_1$ in the linear regression model (6.1.1) and (6.2.1). The

---

9. Note that Kariya and Eaton's (1977) Theorem 3.1, which Corollary 5.1.3 extends, could not have been applied in this case since $w = v/(v'v)^{\frac{1}{2}}$ has a joint density function with respect to the uniform measure on the unit hypersphere $\{w | w \in R^{n-k}, w'w = 1\}$ and not with respect to R^{n-k} as required by Kariya and Eaton.

10. It also appears as Problem 33 in Rao (1973, p.77).

test which rejects H_0 for small values of

$$s = \tilde{u}'\Sigma_1^{-1}\tilde{u}/z'z$$

is UMPI, where $\tilde{u}$ is the vector of GLS residuals assuming covariance matrix Σ_1 and z is the vector of OLS residuals.

Proof: Applying Lemma 6.2 with $T = P_1'$, $U = P_2'$ and $V = \Sigma_1$ and using (6.2.13), we have

$$\begin{aligned} s &= v'(P_1\Sigma_1P_1')^{-1}v/v'v \\ &= u'P_1'(P_1\Sigma_1P_1')^{-1}P_1u/u'P_1'P_1u \\ &= u'(\Sigma_1^{-1}-\Sigma_1^{-1}P_2'(P_2\Sigma_1^{-1}P_2')^{-1}P_2\Sigma_1^{-1})u/u'Mu \\ &= u'(\Sigma_1^{-1}-\Sigma_1^{-1}X(X'\Sigma_1^{-1}X)^{-1}X'\Sigma_1^{-1})u/u'Mu \\ &= \tilde{u}'\Sigma_1^{-1}\tilde{u}/z'z. \end{aligned}$$

The second last equality follows because

$$P_2M = 0$$

implies we can write

$$P_2' = XG,$$

where G is a $k \times k$ non-singular transformation matrix.

Note that s has the same null distribution when u is distributed $N(0,\sigma^2 I_n)$ as it does when u takes any other $E_0(n,I_n)$ distribution. Similarly, for any given Σ_1, s has the same alternative distribution for all $E_0(n,\Sigma_1)$ distributions taken by u, including the $N(0,\sigma^2\Sigma_1)$ distribution.

The s test of Theorem 6.1 is identical to Kadiyala's (1970) test and the "Likelihood Ratio Observable" test constructed by Fraser et al. (1976). The particular form of the test in Corollary 6.1.1 corresponds to Durbin and Watson's (1971) MPI test obtained upon making the corrections to their proof noted in footnote 6.

3. TESTING FOR AUTOCORRELATION USING LUS RESIDUALS

Consider the linear regression model (6.1.1) whose disturbances are generated by the stationary, first-order, autoregressive scheme

$$(6.3.1) \qquad u_t = \rho u_{t-1} + e_t, \qquad |\rho| < 1,$$

with $e = (e_1 \ldots, e_n)'$ being an $E_0(n, I_n)$ random vector. We wish to test the null hypothesis, $H_0: \rho = 0$, against the alternative, $H_a: \rho > 0$, using a test based on LUS residuals. As noted in Section 2 of Chapter 4, u is distributed $E_0(n, \Sigma)$, where Σ is of the form

$$\Sigma = [(1-\rho)^2 I_n + \rho A_1 + \rho(1-\rho)C_1]^{-1},$$

and A_1 and C_1 are defined by (4.2.29) and (4.2.30), respectively.

Theil (1971) has shown that a necessary and sufficient condition for an m-dimensional residual vector to be LUS, is that it can be written in the form

$$(6.3.2) \qquad v = B'y,$$

with B, an $n \times m$ nonstochastic matrix, satisfying

$$(6.3.3) \qquad B'X = 0$$

and

$$(6.3.4) \qquad B'B = I_m .$$

The only assumption made by Theil about the distribution of the disturbances is that they have zero mean and a scalar covariance matrix. Hence this condition applies when the disturbances follow a spherically symmetric distribution with finite second order moments. In view of property III of Chapter 3 and footnote 3, it is also valid for spherically symmetric disturbances for which first order or second order moments do not exist. In addition, Theil showed that m cannot exceed $n - k$.

An alternative necessary and sufficient condition is provided by the following theorem:

Theorem 6.2. A necessary and sufficient condition that an $(n-k)$-dimensional residual vector is LUS is that it can be written in the form (6.3.2), where B is an $n \times (n-k)$ nonstochastic matrix whose columns are the orthonormal characteristic vectors corresponding to the unit roots of the matrix M.

Proof: Necessity. Suppose v is an $(n-k)$-dimensional LUS residual vector written in the form (6.3.2). (6.3.4) implies BB' is idempotent and therefore has $(n-k)$ unit and k zero characteristic values. (6.3.3) and (6.3.4) imply

(6.3.5) $$\begin{bmatrix} B' \\ (X'X)^{-\frac{1}{2}}X' \end{bmatrix} BB'[B \vdots X(X'X)^{-\frac{1}{2}}] = \begin{bmatrix} I_{n-k} & 0 \\ 0' & 0_k \end{bmatrix}$$

and

(6.3.6) $$\begin{bmatrix} B' \\ (X'X)^{-\frac{1}{2}}X' \end{bmatrix} [B \vdots X(X'X)^{-\frac{1}{2}}] = I_n .$$

Hence $$[B \vdots X(X'X)^{-\frac{1}{2}}] \begin{bmatrix} B' \\ (X'X)^{-\frac{1}{2}}X' \end{bmatrix} = I_n ,$$

which implies

$$(6.3.7) \qquad BB' = I_n - X(X'X)^{-1}X'$$
$$= M .$$

Therefore, that the columns of B are orthonormal characteristic vectors of M corresponding to unit roots, follows immediately from (6.3.5) and (6.3.6).

Sufficiency. See Theil (1971, p.208).

A further alternative necessary and sufficient condition is given by the following corollary to Theorem 6.2.

Corollary 6.2.1. Let B_o be any given, $n \times (n-k)$, nonstochastic matrix such that (6.3.3) and (6.3.4) hold. A necessary and sufficient condition that an $(n-k)$-dimensional residual vector, v, is LUS is that it can be written in the form

$$v = G'B_o'y,$$

where G is an $(n-k) \times (n-k)$ orthogonal matrix.

Proof: Necessity. Suppose v is a LUS residual vector. Theorem 6.2 implies it can be written in the form (6.3.2), where B is an $n \times (n-k)$ nonstochastic matrix whose columns form an orthonormal basis of the space spanned by the characteristic vectors corresponding to the unit roots of M. Since the columns of B_o also form such a basis, we can write

$$B = B_oG,$$

where G is an $(n-k) \times (n-k)$ orthogonal matrix.

Sufficiency. Follows because the $n \times (n-k)$ nonstochastic matrix B_oG satisfies (6.3.3) and (6.3.4).

Theorem 6.2 implies that if $v = B'y$ is an $(n-k)$-dimensional, LUS residual vector then the matrix B' satisfies the conditions of the $(n-k) \times n$ matrix, P_1, in Theorem 6.1. Therefore, a UMPI test based on LUS residuals is to reject H_0 for small values of

$$s = v'(B'\Sigma B)^{-1}v/v'v. \tag{6.3.8}$$

Note that for a given design matrix, (6.3.8) is invariant to the choice of B. By Corollary 6.2.1, any B can be written as

$$B = B_oG,$$

where B_o is a fixed $n \times (n-k)$ matrix such that (6.3.3) and (6.3.4) hold and G is an $(n-k) \times (n-k)$ orthogonal matrix. Then

$$\begin{aligned} s &= y'B(B'\Sigma B)^{-1}B'y/y'B'By \\ &= y'B_o(B_o'\Sigma B_o)^{-1}B_o'y/y'B_o'B_oy. \end{aligned}$$

In order to apply this test, we need to know the value of Σ under H_a and the appropriate critical value of s. Since B_0 is a function of X through (6.3.3), s will also be a function of X and, in general, the null distribution of s, and hence critical values of s, will depend on X.

Therefore, it is clear that the use of (6.3.8) as a test statistic defeats the purpose of using LUS residuals when testing for autocorrelation: that is to use a test statistic whose critical values can be tabulated because they are independent of the design matrix. However, we shall see that there are classes of design matrices for which the test based on (6.3.8) can be approximated by a test with a test statistic whose null distribution is independent of X and possible values taken by ρ under H_a.

First, approximate Σ by $\bar{\Sigma}$ as given by (6.2.3). Note that $\bar{\Sigma}$, $\bar{\Sigma}^{-1}$ and A_1 have the same characteristic vectors. As well as being characteristic vectors of the matrix M, suppose the columns of B can also be chosen to correspond to $(n-k)$ characteristic vectors of $\bar{\Sigma}$. Then

$$(B'\bar{\Sigma}B)^{-1} = B'\bar{\Sigma}^{-1}B,$$

and (6.3.8) can be approximated by

$$\begin{aligned}\bar{s}^* &= v'B'\bar{\Sigma}^{-1}Bv/v'v \\ &= (1-\rho)^2 + \rho v'B'A_1Bv/v'v.\end{aligned}$$

Therefore, if the columns of B can be chosen to be $(n-k)$ characteristic vectors of A_1, an approximately UMPI test of H_0 against H_a is to reject H_0 for small values of

$$\begin{aligned}s^* &= v'B'A_1Bv/v'v \\ &= y'MA_1My/y'My,\end{aligned}$$

which is the DW statistic (5.4.9).

From (6.3.3), it follows that the columns of B can be chosen to be $(n-k)$ characteristic vectors of A_1 only when the columns of X can be expressed as a linear combination of the characteristic vectors corresponding to the remaining k characteristic roots of

A_1. Work by Theil and Nagar (1961), Hannan and Terrell (1968)[11] and Dubbelman (1972) suggests that, in economic time series analysis, it is often likely that the space generated by the k columns of X is a good approximation to the space spanned by the characteristic vectors corresponding to the k smallest characteristic roots of A_1.

Denote by

$$\lambda_i = 2\{1 - \cos(\pi(i-1)/n)\}, \qquad i=1,\ldots,n,$$

the n characteristic roots of A_1, let $h_1, \ldots, h_n$ be the corresponding characteristic vectors[12] and let K be the $n \times (n-k)$ matrix

$$K = [h_{k+1},\ldots,h_n].$$

Consider the test statistic,

$$(6.3.9) \qquad r = v'\Lambda v/v'v = \sum_{i=1}^{n-k} \lambda_{k+i} v_i^2 / \sum_{i=1}^{n-k} v_i^2,$$

where Λ is the $(n-k) \times (n-k)$ diagonal matrix with $\lambda_{k+1}, \ldots, \lambda_n$ as its diagonal elements. When the k columns of X can be expressed as linear combinations of $h_1, \ldots, h_k$, r is identical to the DW test statistic and provides an approximately UMPI test of H_0 against H_a.

---

11. Hannan and Terrell showed that the approximation is reasonably good whenever the spectrum of the regressors is relatively concentrated near the origin, as is often the case in economic time series analysis. [See Granger (1966).]

12. Analytic expressions for h_i, $i = 1, \ldots, n$, are given by Abrahamse and Koerts (1971).

If the columns of X can be expressed *approximately* as linear combinations of $h_1, \ldots, h_k$, then rejecting H_0 for small values of (6.3.9) could also be expected to be an approximately UMPI test, especially if the matrix B used to form the LUS residual vector, v, is chosen to approximate K in some optimal manner. There are a number of loss functions that could be used to determine the most appropriate B. For computational simplicity, we choose that B which minimizes the sum of squares of the elements of $(B-K)$. By Corollary 6.2.1, this is equivalent to minimizing

$$\mathrm{tr}\{(B_oG-K)(B_oG-K)'\} = 2(n-k) - 2\mathrm{tr}(B_oGK')$$

with respect to G, where B_o is any given $n \times (n-k)$ matrix such that (6.3.3) and (6.3.4) hold and G is an $(n-k) \times (n-k)$ orthogonal matrix.

The following Lemma, proved by Dubbelman, Abrahamse and Koerts (1972, p.140), provides a solution to this problem.

<u>Lemma 6.3</u>. Suppose H and W are both square matrices of the same order. W is fixed and nonsingular. Then, under the condition

$$H'H = I,$$

maximization of $\mathrm{tr}(HW)$ with respect to H has the unique solution

$$H_0 = W'(WW')^{-\frac{1}{2}}.$$

Since

$$\mathrm{tr}(B_oGK') = \mathrm{tr}(GK'B_o),$$

this lemma can be applied with

$$H = G$$

and

$$W = K'B_o.$$

Assuming $K'B_o$ is nonsingular[13] and making use of (6.3.7), the solution provided by Lemma 6.3 is

$$G_o = B_o'K(K'MK)^{-\frac{1}{2}}.$$

Hence, when the columns of X can be expressed approximately as linear combinations of $h_1, \ldots, h_k$, the LUS residual vector, which is optimal with respect to the arbitrary optimality criterion we have chosen, is

$$v = (K'MK)^{-\frac{1}{2}}K'My.$$

This is our choice of LUS residual vector for use in (6.3.9) when the columns of X can be approximated by linear combinations of $h_1, \ldots, h_k$.

Since

$$\Lambda = K'A_1K,$$

under both H_0 and H_a, we have

$$\begin{aligned} r &= y'BK'A_1KB'y/y'BK'KB'y \\ &= \bar{v}'A_1\bar{v}/\bar{v}'\bar{v}, \end{aligned}$$

---

13. Dubbelman (1973) provides a generalized version of Lemma 6.3 that allows W to be singular, in which case the solution no longer need be unique. We shall restrict our attention to the more interesting case where $K'B_o$, or equivalently $K'MK$, is of full rank.

where

$$\bar{v} = K(K'MK)^{-\frac{1}{2}}K'My$$

is the AK residual vector. Hence, for a given regression equation, (6.3.9) and the AK test statistic will yield the same values. Therefore, our proposed test is equivalent to the AK test. Abrahamse and Koerts (1971, p.74) demonstrated that under H_0, r has the same distribution as the DW upper bound, thus the tabulated values of the DW upper bound provide critical values of (6.3.9).

4. DISCUSSION

The above analysis casts new light on the conclusions drawn by Abrahamse and Koerts (1969, 1971), that the relatively poor performance of the BLUS test with respect to the DW test, when testing for first-order autocorrelation, is caused by the imposed scalar covariance matrix restriction of the BLUS residual vector. Our analysis suggests an alternative explanation - that the poor performance is caused by a combination of the choice of test statistic and the choice of criterion employed to select the particular LUS residual vector to be used in the test. Neither were chosen to maximise the power of the resultant test.

Abrahamse and Koerts suggested relaxing the scalar covariance matrix restriction. They considered LUF (Linear Unbiased with Fixed covariance matrix) residuals which are of the form

$$w = C'y, \tag{6.4.1}$$

where C is an $n \times n$ matrix subject to the restrictions, (6.4.2) $$C'X = 0$$

and

(6.4.3) $$C'C = \Omega,$$

where Ω is an $n \times n$ covariance matrix fixed *a priori*. Ω is required to be idempotent with rank $n - k$, so that one can write

(6.4.4) $$\Omega = LL',$$

where L is an $n \times (n-k)$ matrix such that

(6.4.5) $$L'L = I_{n-k}.$$

Abrahamse and Koerts (1971, p.72) found that any $n \times n$ matrix C, satisfying (6.4.2), (6.4.3), (6.4.4) and (6.4.5), is of the form

$$C = B_o GL',$$

where B_o is an $n \times (n-k)$ matrix of orthonormal characteristic vectors of M corresponding to the unit roots and G is an $(n-k) \times (n-k)$ orthogonal matrix. Theorem 6.2 and Corollary 6.2.1 imply, therefore, that C has the form

$$C = BL',$$

where $B'y$ is an $(n-k)$-dimensional, LUS residual vector. Hence, for each LUF residual vector,

$$w = C'y = LB'y,$$

and each test statistic of the form $f_F(w)$, there is a corresponding LUS residual vector,

$$v = B'y,$$

and a test statistic,

$$f_S(v) = f_F(Lv),$$

such that

$$f_F(w) \equiv f_S(v).$$

That is, for each test based on LUF residuals, there exists a LUS residuals test with the same critical region and which, therefore, is an equivalent test. Clearly, Abrahamse and Koerts could have achieved the same improvement in power of the BLUS test, mentioned in Section 1, using LUS residuals if they had considered changes in the test statistic instead of weakening the scalar covariance assumption.

Theorem 6.1 demonstrates that the UMPI test against an alternative hypothesis that ρ takes a given value, has a critical region dependent on the design matrix. However, any test with a fixed test statistic and based on LUS residuals will have a critical region independent of X. Obviously such a test cannot approximate the UMPI test for all possible design matrices. Our hope for a satisfactory LUS test rests on the assumption that the columns of $n \times k$ design matrices encountered in economic time series analysis can be approximated by linear combinations of the vectors $h_1, \ldots h_k$.

Work by Dubbelman (1972), Dubbelman, Abrahamse and Louter (1978) and Dent and Cassing (1978) indicates that there are a significant proportion of design matrices in an economic context, for which this assumption breaks down with a serious loss of power. In an attempt to find a solution to this problem, Dubbelman (1972) investigated a test procedure which involves a range of tests of the form of the AK test but each with different K matrices, and a selection criterion which has regard to the design matrix, when the choice of test is made. If the columns of a particular X matrix can be approximated by linear combinations of any k characteristic vectors of the matrix A_1, then the analysis of the previous section can be applied to obtain an approximately UMPI test. The matrix K will now be comprised of the remaining $n - k$ characteristic vectors of A_1 and the matrix Λ in (6.3.9) will have the associated characteristic values as its diagonal elements. The selection criterion is to choose K to comprise the $n - k$ h-vectors corresponding to the $n - k$ largest values of $h_i'Mh_i$, $i = 1, 2, \ldots, n$. Whenever the columns of X can be expressed as linear combinations of k of the characteristic vectors of A_1, this selection procedure leads to a test which is identical to the DW test as well as being approximately UMPI. From this, one might expect that for general X, the power of Dubbelman's procedure is reasonably similar to that of the DW test, especially in cases when the AK test performs poorly. This claim tends to be supported by the available empirical evidence.

We can assess Dubbelman's selection criterion in the light of the analysis of Section 3. Using the arguments presented there, an optimal selection criterion would be to choose K to comprise those $n - k$ h-vectors which allow the smallest sum of squares of the elements of $(B-K)$ for the optimal choice of B. This can be shown to be equivalent to choosing a K matrix comprised of h-vectors and which maximizes $\mathrm{tr}\{(K'MK)^{\frac{1}{2}}\}$. Since

$$K'MK = I_{n-k} - K'X(X'X)^{-1}X'K,$$

we can write $(K'MK)^{\frac{1}{2}}$ as the following binomial expansion,[14]

$$(6.4.6) \qquad (K'MK)^{\frac{1}{2}} = I_{n-k} + \sum_{i=1}^{\infty} \frac{-\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{3}{2}\ \cdots\ \cdot(i-\frac{3}{2})}{i!} \{K'X(X'X)^{-1}X'K\}^{i}.$$

Neudecker (1977) demonstrated that all characteristic values of $K'X(X'X)^{-1}X'K$ are non-negative and less than one, thus guaranteeing the validity of the above expansion.

The first two terms of (6.4.6) are

$$\tfrac{1}{2}I_{n-k} + \tfrac{1}{2}K'MK.$$

Since

$$tr(\tfrac{1}{2}I_{n-k}+\tfrac{1}{2}K'MK) = \tfrac{1}{2}(n-k) + \tfrac{1}{2}\sum_{i=1}^{n-k} K_i'MK_i,$$

where K_i, $i = 1, \ldots, n - k$, are the columns of K, Dubbelman's selection criterion is equivalent to choosing the K matrix which maximizes the trace of the first two terms of (6.4.6). Obviously, any selection procedure which uses more than the first two terms of (6.4.6) would be computationally cumbersome. Therefore Dubbelman's selection procedure appears satisfactory in view of the trade-off between accuracy and ease of computation, and his test procedure could be expected to be approximately UMPI whenever the columns of X can be approximated as linear combinations of k characteristic vectors of A_1.

Dubbelman proposes his procedure as a serious alternative to

---

14. See Waugh and Abel (1967). The expansion is valid provided the characteristic roots of $K'X(X'X)^{-1}X'K$ are less than one in absolute value.

the DW test only for small samples,[15] the problem being that for large n and k, an unmanageable number of significance points need to be tabulated. For example, when $n = 60$ and $k = 5$, there are approximately five and a half million possible choices of K, each requiring its own significance points. The advantage of selecting a LUS test from a range of such tests is that it results in a test procedure whose critical region changes to suit the design matrix. The ideal situation would be to identify a handful of such tests that adequately cover the range of design matrices encountered in economic time series analysis.

5. CONCLUSIONS

In this chapter, we have shown that the AK test based on a LUF residual vector is identical to a test using a LUS residual vector. We have also found this test to be approximately UMPI when the regressors can be approximated as linear combinations of the characteristic vectors associated with the k smallest roots of the matrix A_1. Although both methods of applying the test will yield exactly the same results, on reflection perhaps the test using LUS residuals is to be preferred since these residuals have a number of other uses not shared by LUF residuals. For example, LUS residuals can be used to test for spherically symmetric disturbances using any of the numerous tests of independent, identically distributed, normal variates, that were shown to be tests of spherical symmetry in Section 3 of Chapter 5.

---

15. Louter and Dubbelman (1973) tabulated a selection of critical values for design matrices where $9 \leq n \leq 20$ and $2 \leq k \leq 4$.

It would appear that the ideal of a test for autocorrelation, with approximately the same power properties as the DW test for general X, as well as having critical regions independent of the design matrix, does not exist. A possible alternative to this ideal test might be to have a suitable range of tests whose critical regions are independent of X, with the selection of the appropriate test being determined by the form of the design matrix.

CHAPTER 7

THE DURBIN-WATSON BOUNDS TEST AND REGRESSIONS THROUGH THE ORIGIN[1]

1. INTRODUCTION

Let H_ρ denote the linear regression model

$$y = X\beta + u, \tag{7.1.1}$$

where y is an observable, n-dimensional random vector, X is an $n \times k$ nonstochastic matrix of rank $k < n$, β is a k-dimensional vector of unknown parameters and u is an unobservable, n-dimensional disturbance vector whose components are generated by the stationary, first-order, autoregressive scheme,

$$u_t = \rho u_{t-1} + e_t, \qquad |\rho| < 1, \tag{7.1.2}$$

with

$$e = (e_1,\ldots,e_n)'$$

being an $\mathcal{E}_o(n,I_n)$ random vector. Optimal power properties of the DW bounds test of the null hypothesis,

$$H_0: \rho = 0,$$

under the maintained hypothesis H_ρ, were studied in Section 4 of Chapter 5. In this chapter we discuss the application of the DW bounds test to regression equations fitted through the origin.

---

1. A paper based on the material in this chapter was presented to the 12th New Zealand Mathematical Colloquium in May 1977.

Assuming normally distributed disturbances and for regression equations with an intercept term,[2] Durbin and Watson (1951) tabulated[3] bounds for the .05, .025 and .01 significance points of the DW test statistic, d_1, given by (5.4.9), suitable for the test of H_0 against the alternative hypothesis,

$$H_a : \rho > 0.$$

Theorem 5.1 implies these bounds are also valid for $E_o(n, I_n)$ disturbances.

In order to test H_0 with respect to a regression fitted through the origin, Durbin and Watson suggested re-estimating the equation with a superfluous intercept term so that their tabulated bounds can be applied. Recently, Kramer (1971) provided an alternative method of dealing with this case when he published tables of bounds appropriate for testing regressions through the origin against H_a. We shall refer to the former method of applying the DW test as Durbin and Watson's procedure and the use of Kramer's bounds as Kramer's procedure. Note that Theorem 5.1 implies Kramer's bounds are valid for $E_o(n, I_n)$ disturbances.

The following is the layout of this chapter. In the subsequent section, the question of pre-test bias in Kramer's procedure is

---

2. Such regressions include those which do not specify an intercept term but where there exists a linear combination of the regressors that is constant for each observation. Also included are regressions whose regressors and regressand all have zero sample means.

3. These tables have been recomputed by Koerts and Abrahamse (1969) using Imhof's (1961) method of calculating the distribution function of quadratic forms in normal variables. They have been extended to include further values of n and k by Savin and White (1977) and Farebrother (1978a).

discussed. Section 3 is devoted to a comparison of the power properties of the two alternative procedures for selected design matrices. Kramer's procedure is extended in order to test H_0 against negative autocorrelation and tables of the appropriate .05 and .01 significance level bounds are presented in Section 4. Concluding remarks are made in the final section.

2. PRE-TEST BIAS CONSIDERATIONS

One interpretation of the linear regression model is that it is an approximation to a true, and perhaps non-linear, functional relationship. Even were this true functional form to pass through the origin, it need not ensure that the best linear approximation also goes through the origin. Of course, one could use the same reasoning to suggest the inclusion of other variables in the linear regression that are not in the true functional form. On the other hand, that a variable is an argument of the true functional form does not automatically guarantee its place as a regressor in the approximating regression. The question of whether the best linear approximation passes through the origin is no different to the question of whether a particular variable should be included as a regressor in this linear approximation. On balance, it is clear that there are situations in which it is appropriate to fit a linear regression equation through the origin.[4]

---

4. Examples in the econometric literature include regression models fitted by Ando and Modigliani (1963), Almon (1965), Turnovsky and Wachter (1972), Knöbl (1974), Danes (1975) and Hall and King (1976).

Such situations can be classified into two distinct groups:

(i) those in which one is sure that the best linear approximation passes through the origin, and

(ii) those in which one is uncertain as to whether the best linear approximation passes the origin.

Our concern is with the possibility of pre-test bias in tests applied to regressions which fall into the second category.

When the same data are used to help decide whether to fit the regression through the origin, as well as to test for serial correlation in the disturbances, the true size of this latter test will suffer from pre-test bias. This is true whether the final regression specification has been arrived at by a formal test of the null hypothesis of a zero intercept or by a cursory examination of the residuals of the competing specifications, although the degree of bias may differ.

An argument in favour of Durbin and Watson's procedure is that it circumvents this problem of pre-test bias. The practice of always fitting an intercept term before testing for serial correlation ensures that the true significance level agrees with the nominal level, even though the intercept term may often be superfluous. However, applying the DW test in the presence of a superfluous constant dummy variable is not without its cost in terms of the test's power as we shall see from the results of the following section.

3. THE POWERS OF THE ALTERNATIVE PROCEDURES

For any critical value, d_1^*, and any $n \times k$ design matrix, X, the power of the DW test under H_ρ is determined by

$$\Pr(d_1 < d_1^*) = \Pr[u'MA_1Mu/u'Mu < d_1^*]$$

$$= \Pr[u'M(A_1 - d_1^*I_n)Mu < 0],$$

where A_1 is the $n \times n$ matrix (4.2.29) and

$$M = I_n - X(X'X)^{-1}X'.$$

When e is distributed $E_o(n,I_n)$, property III and (7.1.2) imply that the disturbance vector of (7.1.1), u, is distributed $E_o(n,V)$, where

$$V = \begin{bmatrix} 1 & \rho & \rho^2 & \ldots\ldots & \rho^{n-1} \\ \rho & 1 & \rho & & \rho^{n-2} \\ \rho^2 & \rho & 1 & & \rho^{n-3} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \ldots\ldots & 1 \end{bmatrix}.$$

Hence, $\zeta = V^{-\frac{1}{2}}u$ takes an $E_o(n,I_n)$ distribution.

The power of the DW test, therefore, is

(7.3.1.) $$\Pr(d_1 < d_1^*) = \Pr[\zeta'(V^{\frac{1}{2}})'M(A_1 - d_1^*I_n)MV^{\frac{1}{2}}\zeta < 0]$$

$$= \Pr[\zeta'\Gamma\zeta < 0]$$

$$= \Pr[\sum_{i=1}^{n}\gamma_i\zeta_i^2 < 0]$$

$$= E[\chi_\gamma(\zeta)],$$

where $\zeta = (\zeta_1,\ldots,\zeta_n)'$, γ_i, $i = 1, \ldots, n$, are the characteristic roots of

(7.3.2) $$\Gamma = (V^{\frac{1}{2}})'M(A_1 - d_1^*I_n)MV^{\frac{1}{2}},$$

and $\chi_\gamma(\zeta)$ is the indicator function defined by

$$\chi_\gamma(\zeta) = 1 \quad \text{if} \quad \sum_{i=1}^{n} \gamma_i \zeta_i^2 < 0,$$

$$= 0 \quad \text{otherwise.}$$

Note that at least k of the characteristic roots of Γ are zero.

Since $\chi_\gamma(\zeta)$ is invariant to the scale of ζ, Theorem 5.1 implies it has the same distribution whether ζ is assumed to be $E_o(n,I_n)$ or $N(0,\sigma^2 I_n)$. Thus (7.3.1) has the same value for all $E_o(n,I_n)$ distributions followed by ζ including the $N(0,\sigma^2 I_n)$ distribution. When normality is assumed, the subroutine FQUAD, described by Koerts and Abrahamse (1969, p.155) and based on the methods of Imhof (1961), can be used to compute the required probabilities.

A $100\alpha\%$ critical value for the exact DW test against H_a can be obtained by solving

$$\Pr[\sum_{i=1}^{n} \xi_i \zeta_i^2 < 0] = \alpha$$

for d_1^*, where ξ_i, $i = 1, \ldots, n$, are the characteristic roots of $M(A_1 - d_1^* I_n)M$.

The power functions of the DW test using both Kramer's procedure and Durbin and Watson's procedure, against the alternative hypothesis $H_a: \rho > 0$, were evaluated for the underlying regression model, H_ρ with $k = 2$. Three sets of exogenous variables believed to be representative of non-seasonal economic time series were used in the comparison:

(i) The two exogenous variables of Durbin and Watson's (1951, p.159) consumption of spirits example.

(ii) Two independent stationary autoregressive time series generated as

$$x_{it} = .5x_{it-1} + \eta_{it},$$

where

$$\eta_{it} \sim IN(0,1),\ t = -99,\ \ldots,\ 0,\ 1,\ \ldots,\ n,$$

and

$$x_{i,-100} \sim N(0,1.333),^{5}\quad i = 1,\ 2;$$

i.e. two series were generated separately and the first 100 values of each discarded in order to minimize the effects of initializing values.

(iii) Two independent autoregressive time series with linear trend, generated according to

$$x_{it} = z_{it} + .25t,$$

$$z_{it} = .5z_{it-1} + \xi_{it},$$

where

$$\xi_{it} \sim IN(0,1),\ t = -99,\ \ldots,\ 0,\ 1,\ \ldots,\ n,$$

and

$$z_{i,-100} \sim N(0,1.333),\ i = 1,\ 2;$$

i.e. as for (ii), two series were generated separately and the first 100 values of each discarded.

---

5. The choice of variance for the initial value of the series is determined by the variance of the x_i series, $Var(x_i) = 1.333$.

The values of the two sets of artificially generated variables are presented in Appendix 1, while details and properties of the pseudo-random number generator used to generate independent standard normal variates can be found in Appendix 2.

For each of these design matrices, both when a superfluous intercept term is excluded and when it is included, the powers of the DW test using the appropriate five per cent lower bound (d_L), true significance point (d_T) and upper bound (d_U) as critical values were calculated[6] for $n = 15, 60$ and $\rho = 0, .3, .6, .9$.

There are a number of ways in which the DW bounds test can be applied in practice. The two recommended methods are either:

(i) to use the true significance point as the critical value, or

(ii) to use the appropriate bounds in an initial attempt to either accept or reject the null hypothesis, and when the DW statistic falls in the inconclusive region, approximate the true significance point using one of the many procedures reviewed by Durbin and Watson (1971) and Harrison (1972).

We have refrained from calculating powers of the DW test with approximations to the true significance point as critical values. Clearly, one would expect the powers for the better approximations to be close to those for d_T.

The calculated values of the power functions are presented in Table 7.1. The principal values are the powers of the DW test using

---

6. Because the distribution of the DW test statistic under both H_0 and H_a is independent of β and σ^2, these parameters were left undetermined.

Kramer's procedure, while those in parentheses are the corresponding powers for Durbin and Watson's procedure.

Although care must be taken in drawing broad inferences from the limited study made here, several general features emerge. The most striking is that the power of the DW test using Kramer's procedure is never less than that using Durbin and Watson's procedure. This result is not surprising, since the addition of an intercept term causes the loss of a degree of freedom when estimating the disturbance vector, and tests based on these residuals might be expected to be relatively less powerful.

These differences in power are surprisingly large for small sample sizes $(n=15)$ but are of no great consequence for $n = 60$, reflecting the relative importance of the loss of one degree of freedom. The differences tend to be greater when the exogenous series contain trend components. Another interesting feature is that, generally, the probability of the DW statistic lying in the inconclusive region is smaller for Kramer's procedure than for Durbin and Watson's procedure. This is always the case under H_0.

The uniformity of the above inferences for each of the three widely differing design matrices, tends to suggest that they might also apply to a much wider range of design matrices than those considered here.

4. BOUNDS FOR TESTING FOR NEGATIVE AUTOCORRELATION

In this section, we consider the problem of obtaining DW bounds suitable for testing H_0 against the alternative hypothesis of

TABLE 7.1

Powers of the Durbin-Watson Bounds and the Exact Durbin-Watson Test using Kramer's Procedure (and Durbin and Watson's Procedure)

ρ	n = 15 d_L	d_T	d_U	n = 60 d_L	d_T	d_U
	Durbin and Watson's Consumption of Spirits Data					
0	.004 (.002)	.050 (.050)	.069 (.070)	.015 (.014)	.050 (.050)	.051 (.051)
.3	.039 (.023)	.237 (.204)	.289 (.253)	.513 (.488)	.710 (.690)	.711 (.694)
.6	.242 (.125)	.588 (.461)	.642 (.520)	.984 (.977)	.995 (.993)	.995 (.993)
.9	.623 (.290)	.851 (.653)	.876 (.701)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
	Stationary Autoregressive Time Series Data					
0	.005 (.005)	.050 (.050)	.097 (.115)	.020 (.019)	.050 (.050)	.065 (.065)
.3	.049 (.033)	.226 (.187)	.338 (.328)	.560 (.535)	.710 (.691)	.751 (.734)
.6	.263 (.151)	.652 (.436)	.675 (.599)	.989 (.984)	.996 (.994)	.997 (.995)
.9	.675 (.426)	.867 (.700)	.914 (.811)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
	Linear Trend with Autoregressive Disturbances Time Series Data					
0	.004 (.003)	.050 (.050)	.081 (.085)	.017 (.017)	.050 (.050)	.058 (.058)
.3	.042 (.023)	.226 (.187)	.304 (.267)	.537 (.512)	.708 (.689)	.732 (.715)
.6	.242 (.110)	.560 (.406)	.640 (.502)	.987 (.980)	.996 (.993)	.996 (.994)
.9	.630 (.249)	.834 (.562)	.872 (.646)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)

negative autocorrelation,

$$H_a^{\prime}: \rho < 0,$$

when the regression equation in the underlying model, H_ρ, is fitted through the origin. Let $d^0_{L\alpha}$ and $d^0_{U\alpha}$ denote the lower and upper bounds, respectively, of the 100α% DW critical value for a regression through the origin and let $d^1_{L\alpha}$ and $d^1_{U\alpha}$ denote the corresponding bounds for a regression with an intercept term.

Kramer (1971) showed that

$$d^0_{L\alpha} < d^1_{L\alpha}$$

and

$$d^0_{U\alpha} = d^1_{U\alpha}.$$

He also tabulated values of $d^0_{L\alpha}$ with $\alpha = .01, .05$ for testing H_0 against H_a. Durbin and Watson (1951) pointed out that for regressions with an intercept, suitable bounds for tests of H_0 against $H_a^{\prime}$ can be obtained by subtracting tabulated bounds for tests against H_a from the value four; i.e.,

$$d^1_{L(1-\alpha)} = 4 - d^1_{U\alpha}$$

and

$$d^1_{U(1-\alpha)} = 4 - d^1_{L\alpha}.$$

Recent work by King and Giles (1977) on Wallis's (1972) fourth-order analogue of the DW test, implies that these convenient relationships do not hold for regressions without an intercept. However, since

$$d^0_{U\alpha} = d^1_{U\alpha},$$

the upper bound against H'_a is simply the lower bound for regressions with an intercept subtracted from the value four; i.e.,

(7.4.1) $$d^0_{U(1-\alpha)} = 4 - d^1_{L\alpha}.$$

On the other hand, the lower bound, $d^0_{L(1-\alpha)}$, has to be computed using the method outlined in Section 3.

Selected values of $d^0_{L(1-\alpha)}$, for $\alpha = .05, .01$ and $k = 1, \ldots, 6$, are presented in Table 7.2.[7] They were calculated using Koerts and Abrahamse's FQUAD subroutine with maximum integration and truncation errors of 10^{-4}.

Kramer also tabulated values of $d^0_{U\alpha}$ for one regressor equations (k=1) and $\alpha = .05, .01$. These significance points were first tabulated by Koerts and Abrahamse (1969, p.90) as the critical values of Theil's (1965, 1968) BLUS test statistic,

$$Q = \sum_{t=2}^{n} (z_t - z_{t-1})^2 \Big/ \sum_{t=1}^{n} (z_t - \bar{z})^2,$$

where $\bar{z} = \frac{1}{n} \sum_{t=1}^{n} z_t$, and z_t are assumed to be independently, identically, distributed $N(\mu, \sigma^2)$, μ and σ being constants. This follows from the fact that they are also the significance points of d_1 for the regression containing just a constant dummy variable.[8] In this case,

$$d^1_{L\alpha} = d^1_{U\alpha},$$

---

7. Table 7.2 and Kramer's tables of bounds have recently been extended to cover a larger range of values of k and n, in an unpublished paper by Farebrother (1978b).

8. The residuals of such a regression are $y_t - \bar{y}$, $t = 1, \ldots, n$, hence the DW statistic is simply Q.

TABLE 7.2

Calculated Values of $d^{0}_{L(1-\alpha)}$

k	1		2		3		4		5		6	
α	.05	.01	.05	.01	.05	.01	.05	.01	.05	.01	.05	.01
n												
15	2.515	2.818	2.332	2.635	2.126	2.422	1.902	2.185	1.664	1.927	1.419	1.656
16	2.512	2.808	2.344	2.640	2.155	2.447	1.950	2.231	1.732	1.997	1.506	1.749
17	2.508	2.797	2.353	2.643	2.179	2.466	1.990	2.268	1.789	2.054	1.580	1.827
18	2.504	2.787	2.359	2.644	2.199	2.480	2.024	2.299	1.838	2.102	1.644	1.893
19	2.499	2.776	2.364	2.643	2.215	2.492	2.053	2.324	1.880	2.142	1.699	1.948
20	2.494	2.766	2.368	2.641	2.228	2.500	2.077	2.344	1.916	2.176	1.747	1.996
21	2.489	2.756	2.370	2.638	2.239	2.506	2.098	2.361	1.947	2.204	1.789	2.036
22	2.484	2.746	2.372	2.635	2.249	2.511	2.116	2.375	1.974	2.228	1.825	2.071
23	2.479	2.736	2.373	2.631	2.257	2.515	2.131	2.387	1.998	2.249	1.858	2.102
24	2.474	2.727	2.373	2.627	2.263	2.517	2.145	2.396	2.019	2.267	1.886	2.128
25	2.470	2.717	2.373	2.623	2.269	2.518	2.156	2.404	2.037	2.282	1.912	2.151
26	2.465	2.709	2.373	2.618	2.273	2.519	2.167	2.411	2.053	2.295	1.935	2.171
27	2.460	2.700	2.372	2.614	2.277	2.519	2.176	2.416	2.068	2.306	1.955	2.189
28	2.455	2.692	2.371	2.609	2.280	2.519	2.183	2.421	2.081	2.316	1.973	2.205
29	2.451	2.684	2.370	2.604	2.283	2.518	2.190	2.425	2.092	2.325	1.990	2.219
30	2.446	2.676	2.369	2.600	2.285	2.517	2.197	2.428	2.103	2.332	2.004	2.231
31	2.442	2.668	2.367	2.595	2.287	2.515	2.202	2.430	2.112	2.339	2.018	2.242
32	2.438	2.661	2.366	2.590	2.289	2.514	2.207	2.432	2.121	2.344	2.030	2.252
33	2.433	2.654	2.364	2.586	2.290	2.512	2.211	2.433	2.128	2.349	2.041	2.260
34	2.429	2.647	2.362	2.581	2.291	2.510	2.215	2.434	2.135	2.353	2.052	2.268
35	2.425	2.640	2.360	2.576	2.291	2.508	2.218	2.435	2.141	2.357	2.061	2.275
36	2.422	2.634	2.359	2.572	2.292	2.506	2.221	2.435	2.147	2.360	2.069	2.281
37	2.418	2.627	2.357	2.567	2.292	2.503	2.224	2.435	2.152	2.363	2.077	2.287
38	2.414	2.621	2.355	2.563	2.292	2.501	2.226	2.435	2.157	2.365	2.085	2.292
39	2.410	2.615	2.353	2.559	2.292	2.499	2.228	2.435	2.161	2.367	2.091	2.296
40	2.407	2.609	2.351	2.555	2.292	2.496	2.230	2.434	2.165	2.369	2.097	2.300
45	2.391	2.583	2.342	2.535	2.290	2.484	2.236	2.430	2.180	2.374	2.123	2.315
50	2.376	2.559	2.332	2.516	2.287	2.471	2.239	2.424	2.190	2.374	2.139	2.323
55	2.363	2.538	2.324	2.500	2.283	2.459	2.240	2.417	2.196	2.373	2.150	2.327
60	2.351	2.519	2.315	2.484	2.278	2.448	2.240	2.409	2.200	2.370	2.159	2.329
65	2.340	2.503	2.308	2.470	2.274	2.437	2.238	2.402	2.202	2.366	2.165	2.329
70	2.331	2.487	2.300	2.458	2.269	2.427	2.237	2.395	2.203	2.361	2.169	2.327
75	2.322	2.473	2.293	2.446	2.264	2.417	2.235	2.387	2.204	2.357	2.172	2.325
80	2.313	2.461	2.287	2.435	2.260	2.408	2.232	2.380	2.204	2.352	2.174	2.323
85	2.306	2.449	2.281	2.425	2.256	2.399	2.230	2.374	2.203	2.347	2.176	2.320
90	2.299	2.438	2.275	2.415	2.252	2.391	2.227	2.367	2.202	2.342	2.177	2.317
95	2.292	2.428	2.270	2.406	2.248	2.384	2.225	2.361	2.201	2.338	2.177	2.314
100	2.286	2.418	2.265	2.398	2.244	2.377	2.222	2.355	2.200	2.333	2.177	2.310

so that for $k = 1$, Kramer's or Koerts and Abrahamse's tabulated values can be used in relation (7.4.1) to obtain values of $d^0_{U(1-\alpha)}$.

5. CONCLUSIONS

Kramer's bounds for the DW statistic provide a more powerful test for first-order, autoregressive disturbances in regression equations passing through the origin than does Durbin and Watson's suggested procedure of fitting a superfluous intercept and then applying the usual DW bounds test. This is especially true for small sample sizes and indicates the possibility of a serious loss of power in the DW bounds test when few observations are available and irrelevant variables are fitted.

However, if there is doubt that the best linear approximation of the underlying model does in fact pass through the origin, Durbin and Watson's suggestion of fitting an intercept term has merit since the possibility of pre-test bias in the size of the resultant test is avoided. As noted earlier, the use of Durbin and Watson's procedure is not without cost in terms of decreased power.

The decision of which of the two procedures to use should perhaps be based on the degree of one's prior belief that the correct linear model passes through the origin. The results of this chapter suggest the following rule of thumb. If the prior belief is strong, Kramer's procedure should be adopted, otherwise Durbin and Watson's procedure should be used.

When Kramer's procedure is used to test for negative autocorrelation, care must be taken when selecting the appropriate bounds. Table 7.2 together with (7.4.1) provides the correct lower and upper bounds for the DW statistic in this case.

CHAPTER 8

TESTING FOR MOVING AVERAGE DISTURBANCES IN THE LINEAR REGRESSION MODEL

1. INTRODUCTION

In contrast to the vast literature devoted to the subject of testing for AR (Autoregressive) disturbances in the linear regression model, until recently, very little attention had been given to the problem of detecting MA (Moving Average) disturbances. In a survey article on the estimation and use of models with MA disturbances, Nicholls, Pagan and Terrell (1975) attributed the general lack of interest in such models to the computational difficulty involved in estimating their parameters. Fortunately, this problem has been alleviated to some extent by the availability of new numerical algorithms and improvements in computer hardware. Another feature of the contrast between the two sets of literature is that the procedures developed specifically to test for MA disturbances are generally designed for regressions with lagged dependent variables as regressors. Consequently, they are almost all asymptotic tests.

The forerunner[1] of these testing procedures was that proposed by Box and Pierce (1970) for linear autoregressive models without exogenous regressors. Fitts (1973) developed a test for MA(1) (first-order MA) disturbances using Durbin's (1970a) approach to the corresponding problem

1. Obviously, the standard likelihood ratio test can always be applied, although computational difficulties such as inverting the covariance matrix of the disturbance vector make this procedure even more cumbersome than its AR counterpart.

for AR disturbances, while Kenward (1976) presented a test for general ARMA disturbances using the same approach. More recently, Breusch (1978) and Godfrey (1978c) have, independently, derived tests for MA disturbances using the LM (Lagrange Multiplier) test approach of Silvey (1959) and Aitchison and Silvey (1960). Surprisingly, they found that the LM test for MA disturbances of a given order is identical to the LM test for AR disturbances of the same order.

When testing for MA(1) disturbances, this is a notable result for two reasons. First, the LM test is asymptotically equivalent to the likelihood ratio test and, secondly, for nonstochastic regressors, the LM test reduces to the DW test. Therefore, as a test for MA(1) disturbances in the linear regression model (7.1.1), the DW test has desirable large sample power properties.

Although there are sound reasons why MA disturbances are more likely to occur in regression models with lagged dependent variables as regressors,[2] occasionally there is a need to test for MA disturbances in models with nonstochastic regressors. Examples can be found in papers authored by Zellner and Montmarquette (1971), Rowley and Wilton (1973) and Kenward (1975). The purpose of this chapter is to investigate the power properties of the DW test as a test for MA(1) disturbances in such regression models.

The following is a brief outline of this chapter. In Section 2, the DW test is shown to be an approximately locally best invariant test for MA(1) disturbances. An alternative small sample test for MA(1)

---

2. For a discussion of these reasons, see Sims (1974) and Nicholls et al. (1975).

disturbances is constructed in Section 3 and its power properties are compared with those of the DW test, for selected design matrices, in Section 4. The final section contains some concluding remarks.

2. AN OPTIMAL POWER PROPERTY OF THE DW TEST AGAINST MA(1) DISTURBANCES

We are interested in the usual linear regression model, (7.1.1), with disturbance vector u whose components are generated by the stationary MA(1) process,[3]

$$u_t = e_t + \gamma e_{t-1}, \qquad t=1,\ldots,n, \tag{8.2.1}$$

where

$$e = (e_0, e_1, \ldots, e_n)'$$

is an $E_0(n+1, I_{n+1})$ random vector. (8.2.1) and property III of Chapter 3 imply that u is distributed $E_0(n,\Sigma)$, where Σ is the tridiagonal matrix,

$$\Sigma = \begin{bmatrix} 1+\gamma^2 & \gamma & 0 & \cdots\cdots & 0 & 0 \\ \gamma & 1+\gamma^2 & \gamma & & 0 & 0 \\ 0 & \gamma & 1+\gamma^2 & & 0 & 0 \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ 0 & 0 & 0 & & 1+\gamma^2 & \gamma \\ 0 & 0 & 0 & \cdots\cdots & \gamma & 1+\gamma^2 \end{bmatrix} \tag{8.2.2}$$

$$= (1+\gamma)^2 I_n - \gamma A_1 - \gamma C_1,$$

with A_1 and C_1 given by (4.2.29) and (4.2.30), respectively.

---

3. The process is stationary for all values of γ. For it to be invertible, γ must lie in the range $-1 < \gamma < 1$.

Σ can be approximated by

$$\bar{\Sigma} = (1+\gamma)^2 I_n - \gamma A_1, \tag{8.2.3}$$

an approximation which becomes exact as $\gamma \to 0$. Let

$$\Omega(\gamma) = I_n - \{\gamma/(1+\gamma)^2\}A_1,$$

so that

$$\bar{\Sigma} = (1+\gamma)^2 \Omega(\gamma).$$

Consider the problem of testing the null hypothesis,

$$H_0 : \gamma = 0,$$

against the alternative hypothesis,

$$H_a : \gamma > 0.$$

This problem is invariant to transformations of the form

$$y \to \alpha y + X\eta, \tag{8.2.4}$$

where α is a positive scalar and η is a k-dimensional vector. In the following theorem, we prove that for testing H_0 against H_a, the DW test is approximately locally best invariant at $\gamma = 0$. Such a test is one whose power function at $\gamma = 0$ has maximum slope of all invariant tests with *the same size*. Note the distinction between locally best invariant tests and locally MPI tests. The former has a power function with maximum slope at $\gamma = 0$ within the class of similar invariant tests, while the latter has maximum power in the neighbourhood of $\gamma = 0$ within the class of invariant tests.

Recall that Lemma 6.1 states that if a random vector z is $E_0(n,\Sigma)$, then

$$w = z/(z'z)^{\frac{1}{2}} \tag{8.2.5}$$

is also $E_0(n,\Sigma)$. In order to prove the theorem which follows, we need the joint density function of (8.2.5).

Lemma 8.1. If z is an $E_0(n,\Sigma)$ random vector, then the n-dimensional random vector w, defined by (8.2.5), has the joint density function

$$h(w) = \tfrac{1}{2}\Gamma(n/2)\pi^{-n/2}|\Sigma|^{-\frac{1}{2}}(w'\Sigma^{-1}w)^{-n/2} \tag{8.2.6}$$

with respect to the uniform measure on

$$C_n = \{x | x \in R^n,\ x'x = 1\}.$$

Proof: First consider the case where z is an $E_0(n,\Sigma)$ random vector with a joint density function of the form (3.2.7) with respect to the Lebesgue measure on R^n.

Transform z to the n-dimensional spherical coordinates, $(r,\theta_1,\ldots,\theta_{n-1})$, as defined by (3.3.10). The Jacobian[4] of this transformation is

$$r^{n-1}\prod_{k=1}^{n-2}\sin^{n-1-k}\theta_k.$$

Let

$$\begin{cases} w_1(\theta_1,\ldots,\theta_{n-1}) = \cos\theta_1, & \\ w_j(\theta_1,\ldots,\theta_{n-1}) = \left(\prod_{k=1}^{j-1}\sin\theta_k\right)\cos\theta_j, & 2\leq j\leq n-1, \\ w_n(\theta_1,\ldots,\theta_{n-1}) = \prod_{k=1}^{n-1}\sin\theta_k, & \end{cases} \tag{8.2.7}$$

4. See Miller (1964, p.13).

and let w be the n-dimensional vector with components, $w_1(\theta_1,\ldots,\theta_{n-1})$, $\ldots$, $w_n(\theta_1,\ldots,\theta_{n-1})$. The joint density function of z becomes

$$\bar{g}(r,\theta_1,\ldots,\theta_{n-1}) = |\Sigma|^{-\frac{1}{2}}\phi(r^2 w'\Sigma^{-1}w)r^{n-1}\prod_{k=1}^{n-2}\sin^{n-1-k}\theta_k,$$

where dr is with respect to the Lebesgue measure on $[o,\infty)$, $d\theta_j$, $j = 1$, $\ldots$, $n-2$, are with respect to the uniform measure on $[o,\pi]$, and $d\theta_{n-1}$ is with respect to the uniform measure on $[0,2\pi)$. The marginal density function of $(\theta_1,\ldots,\theta_{n-1})$ is

$$\begin{aligned}
\bar{h}(\theta_1,\ldots,\theta_{n-1}) &= \int_o^\infty |\Sigma|^{-\frac{1}{2}}\phi(r^2 w'\Sigma^{-1}w)r^{n-1}\prod_{k=1}^{n-2}\sin^{n-1-k}\theta_k dr \\
&= |\Sigma|^{-\frac{1}{2}}\int_o^\infty \phi(s^2)s^{n-1}ds(w'\Sigma^{-1}w)^{-n/2}\prod_{k=1}^{n-2}\sin^{n-1-k}\theta_k \\
&= \tfrac{1}{2}\Gamma(n/2)\pi^{-n/2}|\Sigma|^{-\frac{1}{2}}(w'\Sigma^{-1}w)^{-n/2}\prod_{k=1}^{n-2}\sin^{n-1-k}\theta_k,
\end{aligned}$$

by a simple change of variables and (3.2.3).

Transforming from $(\theta_1,\ldots,\theta_{n-1})$ to (8.2.5) is straightforward since the components of w are defined by (8.2.7). Hence, when $z \sim E_0(n,\Sigma)$ with a joint density function, w has the joint density function (8.2.6).

Finally, consider the case when z is an $E_0(n,\Sigma)$ random vector without a joint density function. (8.2.5) is invariant to the scale of z. Therefore, Theorem 5.1 implies w has the same distribution for any $E_0(n,\Sigma)$ distribution z may take. In particular, w will have the density function (8.2.6) whenever z is $E_0(n,\Sigma)$ without a joint density function.

<u>Theorem 8.1</u>. For testing $H_0:\gamma = 0$ against $H_a:\gamma > 0$ in the linear regression model (7.1.1) and (8.2.1), where e is distributed $E_0(n+1,I_{n+1})$, the test which rejects H_0 for small values of the DW test statistic, d_1, given by (5.4.9), is an approximately locally best invariant test, where invariance is with respect to transformations of the form (8.2.4).

Proof: Let P be any $n \times n$ matrix such that (6.2.5) and (6.2.6) hold and let P be partitioned as (6.2.7). Then

$$w = P_1 z/(z'P_1'P_1 z)^{\frac{1}{2}}$$

is a maximal invariant, where z is the vector of OLS residuals.[5] Define $m = n - k$. Property III of Chapter 3 implies that the m-dimensional random vector,

$$v = P_1 z = P_1 u,$$

follows an $E_0(m, P_1 \Sigma P_1')$ distribution, where Σ is given by (8.2.2). By Lemma 8.1, w has the joint density function,

$$f(w) = \tfrac{1}{2}\Gamma(m/2)\pi^{-m/2}|P_1 \Sigma P_1'|^{-\frac{1}{2}}\{w'(P_1 \Sigma P_1')^{-1}w\}^{-m/2}$$

with respect to the uniform measure on C_m.

By using (8.2.3) to approximate Σ, this joint density function can be approximated by

$$\begin{aligned}\bar{f}(w) &= \tfrac{1}{2}\Gamma(m/2)\pi^{-m/2}|P_1 \bar{\Sigma} P_1'|^{-\frac{1}{2}}\{w'(P_1 \bar{\Sigma} P_1')^{-1}w\}^{-m/2} \\ &= \tfrac{1}{2}\Gamma(m/2)\pi^{-m/2}|P_1 \Omega(\gamma) P_1'|^{-\frac{1}{2}}\{w'(P_1 \Omega(\gamma) P_1')^{-1}w\}^{-m/2}.\end{aligned}$$

Because w is a maximal invariant, all statistics invariant to transformations of the form of (8.2.4) are functions of w. Therefore, from Ferguson (1967, p.235), an approximately locally best invariant test of H_0 against H_a is given by the critical region,

$$\left.\frac{\partial}{\partial\gamma}\bar{f}(w)\right|_{\gamma=0} > c_1 \left.\bar{f}(w)\right|_{\gamma=0},$$

---

5. See the proof of Theorem 6.1.

where c_1 is an appropriate scalar constant, or equivalently,

(8.2.8) $$\frac{\partial}{\partial\gamma}\bar{f}(w)\Big|_{\gamma=0}\Big/\bar{f}(w)\Big|_{\gamma=0} > c_1.$$

(8.2.9) $$\frac{\partial}{\partial\gamma}\bar{f}(w)\Big|_{\gamma=0}$$

$$= -\tfrac{1}{2}\left[\frac{\partial}{\partial\gamma}\{|P_1\Omega(\gamma)P_1'|\}|P_1\Omega(\gamma)P_1'|^{-1}\bar{f}(w)\right]\Bigg|_{\gamma=0}$$

$$-\tfrac{1}{2}m\left[\frac{\partial}{\partial\gamma}\{w'(P_1\Omega(\gamma)P_1')^{-1}w\}\{w'(P_1\Omega(\gamma)P_1')^{-1}w\}^{-1}\bar{f}(w)\right]\Bigg|_{\gamma=0}$$

and

$$\frac{\partial}{\partial\gamma}\{w'(P_1\Omega(\gamma)P_1')^{-1}w\}\Big|_{\gamma=0}$$

$$= \{-w'(P_1\Omega(\gamma)P_1')^{-1}(\frac{\partial}{\partial\gamma}P_1\Omega(\gamma)P_1')(P_1\Omega(\gamma)P_1')^{-1}w\}\Big|_{\gamma=0}$$

$$= w'P_1A_1P_1'w.$$

$\frac{\partial}{\partial\gamma}\{|P_1\Omega(\gamma)P_1'|\}\Big|_{\gamma=0}$ is a scalar constant which we shall denote by c_2. Therefore, dividing (8.2.9) by $\bar{f}(w)\Big|_{\gamma=0}$ and substituting into (8.2.8) yields the critical region

$$-\tfrac{1}{2}c_2 - \tfrac{1}{2}mw'P_1A_1P_1'w > c_1.$$

Using (6.2.13), this is equivalent to

$$y'MA_1My/y'My < c_3,$$

where c_3 is a scalar constant, thus completing the proof.

Note that Σ can be written as

(8.2.10) $$\Sigma = (1+\gamma)^2 I_n - \gamma A_0,$$

where A_0 is the tridiagonal matrix,

$$A_0 = \begin{bmatrix} 2 & -1 & 0 & \cdots\cdots & 0 & 0 \\ -1 & 2 & -1 & & 0 & 0 \\ 0 & -1 & 2 & & 0 & 0 \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ 0 & 0 & 0 & & 2 & -1 \\ 0 & 0 & 0 & \cdots\cdots & -1 & 2 \end{bmatrix}.$$

Since (8.2.10) expresses Σ in a form similar to that of (8.2.3), there is no need to approximate Σ. The arguments used in the proof of Theorem 8.1 can be used to show that the test which rejects H_0 for small values of

$$d_0 = y'MA_0My/y'My,$$

is locally best invariant. This test is also asymptotically equivalent to the likelihood ratio test in view of its asymptotic equivalence with the DW test.

3. A NEW TEST FOR MA(1) DISTURBANCES

The fact that the DW test is approximately locally best invariant at $\gamma = 0$, against MA(1) disturbances, tells us little about its power when γ lies outside the neighbourhood of $\gamma = 0$. However, we would expect the test to have reasonable power properties for large n because of its asymptotic equivalence to the likelihood ratio test.

The remainder of this chapter is devoted to investigating the power properties of the DW test, when the regression disturbances are generated by the MA(1) process, (8.2.1), with $e \sim \mathcal{E}_0(n+1, I_{n+1})$. These properties have been examined for selected design matrices and assuming normal disturbances by Blattberg (1973) and Smith (1976). Recall that in Chapter 5, it was established that such studies would have yielded identical results under the corresponding assumption of $E_0(n,\Sigma)$ disturbances.

Blattberg compared the power function of the DW test against MA(1) disturbances, with the test's power function against AR(1) disturbances. He concluded that the former power function can often take values greater than those of the latter. Smith compared the performance of the DW test with other tests of the independence of regression disturbances - namely Geary's (1970) sign count test, Durbin's (1969) cumulated periodogram test and Durbin's (1970b) 'exact' alternative to the DW bounds test. Note that none of these tests makes use of the knowledge that the disturbances follow a MA(1) scheme under the alternative hypothesis.

Our study differs from those of Blattberg and Smith in that we compare the power function of the DW test against that of a test constructed using the knowledge that under the alternative hypothesis, the disturbances are generated by a MA(1) process.

One obvious candidate for the choice of such a bench-mark test is the UMPI test of Theorem 6.1. The main attraction of this test is that it would provide an upper bound for the power function of any invariant test for MA(1) disturbances. A major drawback is that the value taken by γ under H_a needs to be known in order to apply the test. Thus, it cannot be regarded as a practical test for our particular problem. For this reason, and also because of the high computational cost involved in calculating its power, it was decided not to use this test. Instead, a comparatively simple test was constructed to fill the role of bench-mark test.

The theoretical basis of this test follows from an observation made by Pesaran (1973). He noted that when the disturbances are generated by the MA(1) process, (8.2.1), the eigenvalues of the disturbance vector's characteristic matrix, (8.2.2), are

$$\lambda_i = \gamma^2 + 2\gamma\cos(i\pi/(n+1)) + 1, \qquad i=1,\dots,n,$$

and the associated orthogonal eigenvectors are

$$\tau_i = (2/(n+1))^{\frac{1}{2}}(\sin(i\pi/(n+1)),\sin(2i\pi/(n+1)),\ \dots,\ \sin(ni\pi/(n+1)))', \qquad i=1,\dots,n.$$

Let $T = (\tau_1,\tau_2,\dots,\tau_n)$.[6] Then,

$$T\Sigma T = \Lambda,$$

where Λ is a diagonal matrix with the eigenvalues, λ_j, $j = 1, \dots, n$, as its non-zero elements. Transforming (7.1.1) by premultiplying by T yields

$$\bar{y} = \bar{X}\beta + \bar{u}, \tag{8.3.1}$$

where $\bar{y} = Ty$, $\bar{X} = TX$ and $\bar{u} = Tu$. Property III of Chapter 3 implies $\bar{u}$ is distributed $E_0(n,\Lambda)$.

Therefore, if u is normally distributed with covariance matrix $\sigma^2\Sigma$, where Σ is defined by (8.2.2), then the disturbances of (8.3.1) will have variances,

---

6. Note that T is a symmetric matrix.

$$\bar{\sigma}_t^2 = \sigma^2(\gamma^2+2\gamma\cos(\pi t/(n+1))+1), \qquad t=1,\ldots,n.$$

Clearly, they are homoscedastic if and only if $\gamma = 0$; i.e., if and only if H_0 holds. A plot of these variances against time, for positive γ, is presented in Figure 8.1. The shape of this curve, especially its comparative flatness near $t = 1$ and $t = n$, suggests that Goldfeld and Quandt's (1965) F-test for heteroscedasticity, applied to the transformed regression model (8.3.1), might provide a test for MA(1) disturbances in (7.1.1) with reasonably good power properties. As noted in Chapter 5, Corollary 5.1.1 implies that this test has the same size for all $E_0(n,I_n)$ distributions followed by $\bar{u}$, and for any particular Λ, the same power for all $E_0(n,\Lambda)$ distributions taken by $\bar{u}$.

In summary, Goldfeld and Quandt's F-test for heteroscedasticity, applied to (8.3.1), provides a test of $H_0{:}\gamma = 0$, against $H_a{:}\gamma > 0$, with respect to the linear regression model (7.1.1) and (8.2.1) when e is an $E_0(n+1,I_{n+1})$ random vector. Both the size and, for any fixed value of γ under H_a, the power of the test are invariant to the type of $E_0(n+1,I_{n+1})$ distribution e takes.

The mechanics of the test are as follows: Separate regressions are fitted to the first m and the final m transformed observations, where $k < m \leqslant {}^n/_2$. The ratio of the sum of squared residuals of the first regression to the sum of squared residuals of the second regression is tested against the central F distribution with $(m-k,m-k)$ degrees of freedom.

When $2m < n$, let $\bar{y}$, $\bar{X}$ and $\bar{u}$ be partitioned as

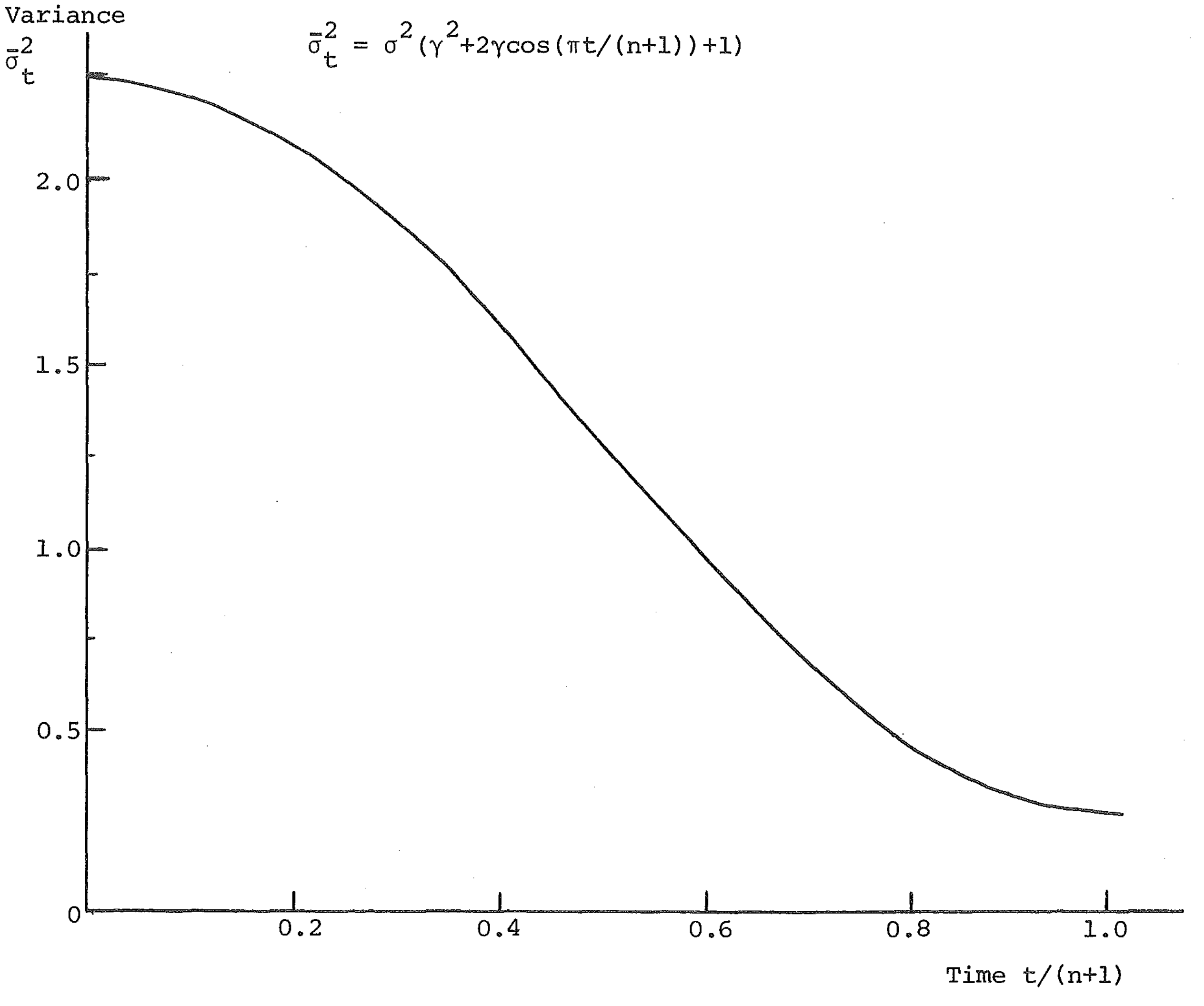
$$\bar{y} = \begin{bmatrix} \bar{y}_{(1)} \\ \bar{y}_{(2)} \\ \bar{y}_{(3)} \end{bmatrix}, \quad \bar{X} = \begin{bmatrix} \bar{X}_{(1)} \\ \bar{X}_{(2)} \\ \bar{X}_{(3)} \end{bmatrix} \quad \text{and} \quad \bar{u} = \begin{bmatrix} \bar{u}_{(1)} \\ \bar{u}_{(2)} \\ \bar{u}_{(3)} \end{bmatrix},$$

where $\bar{y}_{(1)}$, $\bar{y}_{(3)}$, $\bar{u}_{(1)}$ and $\bar{u}_{(3)}$ are m-dimensional vectors, $\bar{y}_{(2)}$ and $\bar{u}_{(2)}$ are (n-2m)-dimensional vectors, $\bar{X}_{(1)}$ and $\bar{X}_{(3)}$ are $m \times k$ matrices and $\bar{X}_{(2)}$ is $(n-2m) \times k$. In the case when $n = 2m$, partition $\bar{y}$, $\bar{X}$ and $\bar{u}$ as

$$\bar{y} = \begin{bmatrix} \bar{y}_{(1)} \\ \bar{y}_{(3)} \end{bmatrix}, \quad \bar{X} = \begin{bmatrix} \bar{X}_{(1)} \\ \bar{X}_{(3)} \end{bmatrix} \quad \text{and} \quad \bar{u} = \begin{bmatrix} \bar{u}_{(1)} \\ \bar{u}_{(3)} \end{bmatrix},$$

FIGURE 8.1

The Variance of the Transformed Disturbances as a Function of Time when the Underlying Disturbances Follow a MA(1) Scheme with $\gamma = .5$ and $\sigma^2 = 1$.



where $\bar{y}_{(i)}$, $\bar{X}_{(i)}$, and $\bar{u}_{(i)}$, $i = 1, 3$, have the same dimensions as above. Let

$$\bar{M}_{(i)} = [I_m - \bar{X}_{(i)}(\bar{X}'_{(i)}\bar{X}_{(i)})^{-1}\bar{X}'_{(i)}], \qquad i=1,3.$$

The suggested test of $H_0:\gamma = 0$, against $H_a:\gamma > 0$, is to reject H_0 when

$$\bar{F} = \frac{\bar{y}'_{(1)}\bar{M}_{(1)}\bar{y}_{(1)}}{\bar{y}'_{(3)}\bar{M}_{(3)}\bar{y}_{(3)}} > F_\alpha(m-k,m-k),$$

where $F_\alpha(m-k,m-k)$ is the $100(1-\alpha)\%$ significance point of the central F distribution with $(m-k,m-k)$ degrees of freedom.

For any critical value F^*, the power of this test is given by

$$\text{(8.3.2)} \qquad \Pr(\bar{F} > F^*) = \Pr[-\bar{u}'_{(1)}\bar{M}_{(1)}\bar{u}_{(1)} + F^*\bar{u}'_{(3)}\bar{M}_{(3)}\bar{u}_{(3)} < 0].$$

Let $\bar{v}$ be the 2m-dimensional vector,

$$\bar{v} = \begin{bmatrix} \bar{u}_{(1)} \\ \bar{u}_{(3)} \end{bmatrix}.$$

Property III of Chapter 3 implies $\bar{v} \sim E_0(2m,\Lambda(m))$, where $\Lambda(m)$ is the $2m \times 2m$ diagonal matrix,

$$\Lambda(m) = \begin{bmatrix} \lambda_1 & & & & & 0 \\ & \ddots & & & & \\ & & \lambda_m & & & \\ & & & \lambda_{n-m+1} & & \\ & & & & \ddots & \\ 0 & & & & & \lambda_n \end{bmatrix}.$$

Let $N(F^*)$ denote the $2m \times 2m$ matrix,

$$N(F^*) = \left[\begin{array}{c:c} -\bar{M}_{(1)} & 0 \\ \hdashline 0 & F^*\bar{M}_{(3)} \end{array}\right].$$

The power function, (8.3.2), can be written as

$$(8.3.3) \qquad \Pr(\bar{F} > F^*) = \Pr[\bar{v}'N(F^*)\bar{v} < 0]$$
$$= \Pr[\zeta'(\Lambda^{\frac{1}{2}}(m))'N(F^*)\Lambda^{\frac{1}{2}}(m)\zeta < 0]$$
$$= \Pr[\zeta'\Gamma\zeta < 0]$$
$$= \Pr[\sum_{i=1}^{n} \xi_i\zeta_i^2 < 0],$$

where $\zeta = (\zeta_1,\ldots,\zeta_n)'$ is distributed $E_0(n,I_n)$ and ξ_i, $i = 1, \ldots, n$, are the characteristic roots of

$$\Gamma = (\Lambda^{\frac{1}{2}}(m))'N(F^*)\Lambda^{\frac{1}{2}}(m),$$

at least $2k$ of which are zero. For the reasons described in Section 3 of Chapter 7, particular values of the power function (8.3.3) can be calculated using Koerts and Abrahamse's (1969, p.155) FQUAD subroutine.

Note that for any critical value, d_1^*, the power function of the DW test, against the alternative of MA(1) disturbances, is given by (7.3.1), where in (7.3.2), V is replaced by Σ. It also may be calculated using the FQUAD subroutine.

4. A COMPARISON OF POWER PROPERTIES OF THE TWO TESTS

Values of the power functions of the DW test and the F test were evaluated for the regression model (7.1.1) and (8.2.1), with $k = 3$ and $e \sim E_0(n+1,I_{n+1})$, under the alternative hypothesis, $H_a:\gamma > 0$. The three sets of exogenous variables described in Chapter 7, with a constant dummy variable included in each set, were used in the comparison. For each of these design matrices, the power function of the DW test, with, alternatively, the appropriate five percent lower bound (d_L), true significance level (d_L) and upper bound (d_U) as critical values, was evaluated[7] for $n = 15, 30, 60$ and $\gamma = 0.0, 0.25, 0.5, 0.75, 0.9999$.

Before the F test can be applied, a value for m must be chosen. In order to determine the effect of this choice, the power of the F test at the five percent significance level was calculated for an exhaustive range of values of m, for each of the design matrices and each of the values of γ employed for the DW test. The values of m used were $m = 5, 6, 7$ for $n = 15$; $m = 6, \ldots, 15$ for $n = 30$; and $m = 6, \ldots, 30$ for $n = 60$.

In certain cases, especially for small degrees of freedom, the true size of the F test was found to differ from the nominal significance level by more than .1%. This occurred when the appropriate values of F^* were taken from the usual tables of the central F distribution critical points[8] and despite the fact that the maximum combined truncation and integration errors in the FQUAD subroutine had been set at .02%. In order to guard against the possibility of inaccurate critical values jeopardizing a fair comparison of powers, the required significance points were calculated using the FQUAD subroutine. This was done by solving

$$\Pr\left[\sum_{i=1}^{m} \zeta_i^2 - F^*(m,m) \sum_{i=m+1}^{2m} \zeta_i^2 < 0\right] = .05$$

---

7. Both the DW test statistic and the F test statistic, and hence their calculated powers, are invariant to β and non-zero values of σ^2.

8. Significance points of the central F distribution were first computed by Merrington and Thompson (1943) on a hand calculator using a combination of eight different numerical techniques. It would appear that almost all other published tables of central F distribution critical values are reproductions of these original tables.

for $F^*(m,m)$, where

$$\zeta = (\zeta_1, \ldots, \zeta_{2m})'$$

is distributed $N(0, I_{2m})$. The computed values of $F^*(m,m)$ for $m = 2, \ldots, 50$ are presented in Table 8.1.

The results of the power calculations can be found in Tables 8.2-8.4. The affect of using the critical values tabulated in Table 8.1 in place of Merrington and Thompson's (1943) significance points was minimal; only a handful of F test powers changed by .001. Our earlier fears of an unfair comparison of powers appear to have been without foundation.

As one might expect, the power of all versions of the DW and F tests increase with sample size as well as with the value of γ, other things being equal. A more interesting feature is that the power of the DW exact test is always greater than that of any version of the F test for $\gamma = .25$ and $.5$. On the other hand, when $\gamma = .75$ and .9999, with one exception,[9] there is always a value of m for which the power of the F test is greater than that of the DW exact test.

The effect of the choice of m value on the power of the F test is not particularly straightforward. For $n = 15$, $m = 7$ always gives the greatest power, while for $n = 30$ and $n = 60$, the value of m giving maximum power shows a tendency to decrease as γ increases. For example, when $n = 60$ and for all three design matrices, $m = 24$ for $\gamma = .25$; $m = 21$ for $\gamma = .5$; $m = 16$ for $\gamma = .75$; and $m = 13$

---

9. The linear trend with autoregressive disturbances time series data for $n = 15$.

TABLE 8.1

Five Percent Critical Values of the One-tailed F Test

Degrees of Freedom	Critical Value	Degrees of Freedom	Critical Value
		26, 26	1.9292
2, 2	19.003	27, 27	1.9048
3, 3	9.281	28, 28	1.8821
4, 4	6.392	29, 29	1.8608
5, 5	5.0534	30, 30	1.8409
6, 6	4.2860	31, 31	1.8221
7, 7	3.7885	32, 32	1.8045
8, 8	3,4390	33, 33	1.7878
9, 9	3.1795	34, 34	1.7721
10, 10	2.9786	35, 35	1.7571
11, 11	2.8181	36, 36	1.7430
12, 12	2.6867	37, 37	1.7295
13, 13	2.5770	38, 38	1.7167
14, 14	2.4838	39, 39	1.7045
15, 15	2.4035	40, 40	1.6928
16, 16	2.3335	41, 41	1.6816
17, 17	2.2719	42, 42	1.6710
18, 18	2.2172	43, 43	1.6607
19, 19	2.1683	44, 44	1.6509
20, 20	2.1242	45, 45	1.6415
21, 21	2.0842	46, 46	1.6325
22, 22	2.0478	47, 47	1.6238
23, 23	2.0144	48, 48	1.6154
24, 24	1.9838	49, 49	1.6073
25, 25	1.9555	50, 50	1.5995

TABLE 8.2

Powers of the DW Test Using Alternative Critical Values and

Powers of the F Test Using Alternative Values of m for n = 15

γ	DW Test			F Test		
	d_L	d_T	d_U	m = 7	m = 6	m = 5
	Durbin and Watson's Consumption of Spirits Data					
0.0	.002	.050	.070	.050	.050	.050
0.25	.014	.165	.212	.144	.123	.105
0.5	.047	.349	.418	.329	.258	.209
0.75	.091	.494	.569	.512	.399	.342
0.9999	.108	.538	.613	.572	.451	.403
	Stationary Autoregressive Time Series Data					
0.0	.005	.050	.115	.050	.050	.050
0.25	.021	.149	.282	.133	.117	.106
0.5	.057	.297	.481	.278	.234	.216
0.75	.094	.410	.609	.411	.353	.363
0.9999	.108	.444	.644	.455	.397	.432
	Linear Trend with Autoregressive Disturbances Time Series Data					
0.0	.003	.050	.085	.050	.050	.050
0.25	.014	.152	.228	.131	.118	.107
0.5	.042	.305	.415	.273	.238	.219
0.75	.073	.422	.544	.404	.362	.372
0.9999	.085	.458	.580	.447	.407	.447

TABLE 8.3

Powers of the DW Test Using Alternative Critical Values and

Powers of the F Test Using Alternative Values of m for n = 30

γ	DW Test			F Test									
	d_L	d_T	d_U	m = 15	m = 14	m = 13	m = 12	m = 11	m = 10	m = 9	m = 8	m = 7	m = 6
	Durbin and Watson's Consumption of Spirits Data												
0.0	.008	.050	.056	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
0.25	.092	.303	.325	.268	.276	.280	.279	.270	.261	.242	.224	.195	.159
0.5	.345	.675	.698	.611	.640	.660	.671	.659	.656	.619	.591	.530	.418
0.75	.565	.851	.866	.787	.822	.850	.869	.868	.882	.863	.867	.849	.746
0.9999	.626	.886	.898	.824	.858	.886	.906	.907	.924	.912	.925	.930	.866
	Stationary Autoregressive Time Series Data												
0.0	.014	.050	.089	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
0.25	.134	.296	.409	.263	.268	.268	.270	.258	.250	.234	.216	.189	.159
0.5	.418	.652	.762	.591	.610	.618	.638	.613	.609	.581	.550	.486	.405
0.75	.628	.827	.899	.762	.787	.803	.834	.816	.828	.814	.807	.763	.706
0.9999	.682	.863	.923	.799	.825	.841	.873	.859	.874	.866	.869	.842	.816
	Linear Trend with Autoregressive Disturbances Time Series Data												
0.0	.011	.050	.074	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
0.25	.113	.295	.370	.263	.267	.268	.270	.261	.252	.236	.214	.188	.157
0.5	.377	.653	.730	.593	.610	.621	.640	.631	.620	.595	.545	.482	.399
0.75	.589	.829	.880	.766	.789	.806	.837	.839	.844	.836	.803	.759	.694
0.9999	.645	.865	.908	.804	.827	.845	.876	.882	.890	.889	.867	.839	.804

TABLE 8.4

Powers of the DW Test Using Alternative Critical Values and

Powers of the F Test Using Alternative Values of m for n = 60

γ	DW Test			F Test									
	d_L	d_T	d_U	m = 30	m = 29	m = 28	m = 27	m = 26	m = 25	m = 24	m = 23	m = 22	m = 21
	Durbin and Watson's Consumption of Spirits Data												
0.0	.014	.050	.051	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
0.25	.318	.536	.541	.476	.486	.493	.499	.506	.510	.512	.510	.507	.504
0.5	.834	.942	.943	.890	.901	.910	.917	.925	.931	.936	.936	.937	.939
0.75	.966	.993	.993	.971	.977	.981	.984	.987	.990	.992	.992	.993	.994
0.9999	.980	.996	.996	.980	.985	.988	.990	.993	.995	.996	.996	.997	.997
	Stationary Autoregressive Time Series Data												
0.0	.019	.050	.065	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
0.25	.358	.530	.584	.466	.476	.484	.490	.496	.500	.503	.501	.501	.499
0.5	.854	.935	.952	.877	.888	.899	.906	.914	.920	.926	.927	.931	.932
0.75	.970	.991	.994	.964	.970	.976	.979	.983	.986	.989	.990	.991	.992
0.9999	.982	.995	.997	.975	.980	.984	.987	.990	.992	.994	.994	.996	.996
	Linear Trend with Autoregressive Disturbances Time Series Data												
0.0	.017	.050	.059	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
0.25	.338	.532	.564	.466	.476	.483	.490	.497	.501	.503	.502	.503	.501
0.5	.844	.937	.947	.878	.890	.899	.906	.915	.922	.927	.928	.933	.935
0.75	.967	.992	.993	.965	.971	.976	.980	.984	.987	.989	.990	.992	.993
0.9999	.981	.996	.997	.976	.981	.984	.987	.990	.992	.994	.995	.996	.997

TABLE 8.4 (continued)

Powers of the DW Test Using Alternative Critical Values and

Powers of the F Test Using Alternative Values of m for n = 60

γ	F Test														
	m = 20	m = 19	m = 18	m = 17	m = 16	m = 15	m = 14	m = 13	m = 12	m = 11	m = 10	m = 9	m = 8	m = 7	m = 6
	Durbin and Watson's Consumption of Spirits Data														
0.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
0.25	.498	.488	.478	.466	.454	.438	.422	.401	.377	.350	.320	.286	.251	.212	.170
0.5	.938	.933	.931	.925	.922	.914	.905	.891	.871	.844	.807	.753	.683	.590	.468
0.75	.994	.994	.994	.994	.995	.994	.995	.994	.992	.990	.986	.976	.959	.925	.852
0.9999	.998	.998	.998	.998	.998	.998	.999	.999	.998	.998	.998	.996	.993	.986	.972
	Stationary Autoregressive Time Series Data														
0.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
0.25	.493	.485	.478	.466	.455	.439	.421	.400	.377	.348	.320	.286	.251	.212	.170
0.5	.931	.930	.929	.924	.920	.912	.903	.887	.870	.837	.802	.750	.683	.589	.469
0.75	.992	.993	.993	.994	.994	.994	.994	.993	.992	.987	.984	.974	.959	.922	.854
0.9999	.997	.997	.998	.998	.998	.998	.998	.998	.998	.997	.997	.995	.992	.985	.974
	Linear Trend with Autoregressive Disturbances Time Series Data														
0.0	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
0.25	.494	.487	.479	.467	.455	.439	.420	.399	.377	.347	.319	.286	.251	.212	.170
0.5	.934	.933	.931	.926	.922	.914	.902	.887	.870	.836	.803	.751	.685	.592	.469
0.75	.993	.994	.994	.994	.995	.995	.994	.993	.992	.987	.984	.975	.960	.927	.854
0.9999	.997	.997	.998	.998	.998	.998	.998	.998	.998	.997	.997	.995	.993	.988	.973

or 14 for $\gamma = .9999$ yield the F test with maximum power. Another feature is that for $n = 30$, power as a function of m often has two or three local maxima, other things remaining equal. This property is not particularly evident for $n = 60$, perhaps suggesting that the power functions of the F test become more well-behaved as n increases.

As a rule of thumb, Goldfeld and Quandt (1965) suggested choosing $m = {}^{n}/_{3}$. Our results indicate that this rule provides an F test with comparatively good power properties for $n = 30$ and $n = 60$, but not for $n = 15$. Because the value of m yielding maximum power is a function of γ, it is difficult to give a universal rule for choosing an optimal value of m. However, Goldfeld and Quandt's rule of thumb, modified so that m is chosen to be greater than ${}^{n}/_{3}$ when $n < 30$, does appear as though it would provide an F test with comparatively favourable power properties.

The limited nature of the above study means that caution should be exercised in any attempt to draw broad inferences from the results. At least for the particular design matrices used, the DW exact test has reasonably good power properties against MA(1) disturbances, although the powers for $n = 15$ should perhaps only be described as fair. It also appears that as γ increases, the DW exact test becomes less competitive in terms of power. The results discourage the use of the DW test with the DW lower bound as a critical value, especially for small samples. If the true critical value is not readily available, it should be approximated by one of many methods reviewed by Durbin and Watson (1971) and Harrison (1972).

The F test, with the above rule of thumb for determining m, is a worthwhile alternative to the DW exact test. Transforming the data[10] and fitting two regressions does involve a reasonable amount of computation; clearly less than is required to establish the true significance point of the DW statistic, but possibly more than would be needed to approximate the DW critical value. On the other hand, this latter alternative provides only an approximate test while the F test is exact.

5. CONCLUSIONS

In this chapter, the DW exact test was shown to possess some desirable power properties when used as a test for MA(1) disturbances in the linear regression model. It was found to be approximately locally best invariant at $\gamma = 0$, while for a selected number of design matrices, it was shown to have good power for larger values of γ and $n \geqslant 30$.

The limited comparison of the power functions of the DW exact test and the F test reported above, suggests that the latter test may provide a useful alternative to the DW exact test, especially when the true significance point for this test is not readily available. Before the F test can be applied, a value for m needs to be chosen. The evidence presented in this chapter suggests that comparatively favourable power properties might result from setting $m = {}^{n}/_{3}$ for $n \geqslant 30$, and for $n < 30$, choosing a value of m closer to ${}^{n}/_{2}$.

---

10. Not all the data need be transformed, only the first m and last m observations.

CHAPTER 9

SUMMARY AND CONCLUSIONS

In this thesis, a number of aspects of the theory of statistical inference associated with the linear regression model were investigated. The principal aim was to examine the implications of replacing the usual normality assumption with the broader assumption that the regression disturbances follow an elliptically symmetric distribution. A further objective was to analyse different features of the problem of testing for serial correlation in elliptically symmetric disturbances, giving emphasis to the power properties of test procedures.

In Chapter 2, the conventional Haavelmo (1944) justification for the normality assumption in econometrics was critically reviewed. This, together with an examination of some of the available empirical evidence, provided the motivation for wanting to consider a theory of statistical inference based on a wider class of disturbance distributions.

The main conclusion that arose from the investigation of the theory of statistical inference for elliptically symmetric regression disturbances reported in Chapters 4 and 5, was that little changes from a practical point of view when the classical normality assumption is replaced by one of elliptical symmetry.

Assuming elliptically symmetric disturbances, necessary and sufficient conditions were found for the weak consistency of any linear "unbiased" estimator of β and for the strong consistency of the OLS and the GLS estimators of β in the usual linear regression model, (7.1.1). The GLS estimator was shown to satisfy a stringent, but natural, optimality property, while for a slightly restricted class of elliptically symmetric disturbance distributions, it was found to be the maximum likelihood estimator.

Chapter 5 considered the small sample properties of statistical tests and estimators associated with the linear regression model when the disturbances are elliptically symmetric. Statistics, which are invariant to the scale of the disturbances, were found to have the same small sample distributions as they do under normality. This implies that the following are invariant to the broadening of the normality assumption to one of elliptical symmetry:

(i) the size of tests based on statistics which are invariant to the scale of the disturbances;

(ii) the power of tests for serial correlation and heteroscedasticity, based on statistics which are invariant to the scale of the disturbances;

(iii) the optimal power properties of particular tests for heteroscedasticity or serial correlation which hold within the class of tests invariant to the scale of the disturbances;

(iv) the small sample distributions of nuisance parameter estimators which are invariant to the scale of the disturbances.

The other important result established in Chapter 5 concerns the distribution of any arbitrary statistic associated with the linear regression model, (7.1.1), under a slightly restricted version of the elliptical symmetry assumption. We found that such a distribution can be viewed as a weighted average of the distributions taken by the statistic for different scales of the disturbances assuming normality; the weights depending on the particular elliptically symmetric distribution the disturbances follow.

In contrast, the standard test of $H_0: \sigma^2 = \sigma_0^2$ assuming $u \sim N(0, \sigma^2 I_n)$ and based on the chi-squared distribution, where σ_0^2 is a fixed scalar, breaks down when the assumption of normality is broadened to one of spherical symmetry with finite second order moments. Also, the maximum likelihood estimator of σ^2 is not invariant to the type of spherically symmetric distribution the disturbances follow. Asymptotic normality of the GLS and the OLS estimators of β in (7.1.1), is another property not invariant to the broadening of the assumption of normally distributed disturbances. On the other hand, those asymptotic tests associated with (7.1.1), whose critical regions are invariant to the scale of the disturbances, remain valid asymptotic tests for a slightly restricted class of elliptically symmetric disturbance distributions.

It therefore appears that from a practical point of view, there are few changes of major consequence in the distributions and properties of estimators and the properties of statistical tests associated with the linear regression model when the assumption of elliptically symmetric disturbances replaces one of normality.

To what extent does this generalization answer the objections to the normality assumption raised in Chapter 2?

One reason for considering the class of elliptically symmetric disturbance distributions is that a large subclass of these distributions have univariate marginal distributions with probability density functions of the form of (2.2.1), thus rebutting one of the theoretical criticisms of the normality assumption raised in Chapter 2. However, elliptically symmetric distributions do possess properties which mitigate against their being able to answer satisfactorily some of the other objections to normality.

That the only elliptically symmetric random vectors with independent components are those having a normal distribution, might be considered at first sight to be one such property. Certainly in the case where the disturbances are known to be independent from observation to observation, elliptical symmetry is not a generalization of normality. However, as noted in Chapter 2, the assumption of independent disturbances is not particularly realistic for the majority of economic applications. Lack of evidence of correlation between disturbance terms is not in itself proof of independence, except for normally distributed disturbances. There is always the possibility of uncorrelated but dependent disturbances, which is the case for non-normal spherically symmetric disturbances whose second order moments exist.

A more serious limitation is the special form of dependency imposed by the elliptical symmetry assumption. This is particularly evident from the property revealed in Chapter 5, that tests for normality with unknown variance are effectively tests for elliptical symmetry. This implies that given one observation of a random vector whose variance is unknown, it is impossible to distinguish between a normal and a non-normal elliptically symmetric parent distribution. Hence, the empirical evidence against normality reviewed in Chapter 2, is, in reality, as just as valid evidence against elliptically symmetry.

In summary, we must therefore conclude that the generalization is of less value than might appear at first sight since it seems to answer only one of a number of theoretical objections that can be made against the normality assumption and none of the empirical objections. The problem inherent in the normality assumption has been found to be, in the large part, identical to that of the elliptical symmetry assumption.

The second half of the thesis reported three contributions on the subject of testing for serial correlation in regression disturbances.

In Chapter 6, an attempt was made to find an "optimal" exact test for first-order autoregressive disturbances based on LUS residuals. The attempt largely failed, although a UMPI test was found for the special case when the characteristic matrix of the disturbances under the alternative hypothesis, is known. For the general problem, a LUS test which is optimal when the column space of the design matrix is a good approximation to the space spanned by the characteristic vectors corresponding to the k smallest characteristic roots of the first difference matrix, A_1, was constructed. This test was found to be identical to the Abrahamse-Koerts test based on a LUF residual vector.

Durbin and Watson's and Kramer's procedures for applying the DW bounds test to a regression equation without an intercept, were compared in Chapter 7. The use of Kramer's bounds was found to provide a more powerful test for first-order autoregressive disturbances than does Durbin and Watson's suggested procedure of fitting a superfluous intercept and then applying the usual DW bounds test. In the case where there is doubt that the best course is to fit the regression equation through the origin, Kramer's procedure will suffer from pre-test bias if the sample data are used to help resolve this doubt. The following rule of thumb was suggested. If the prior belief that the correct linear model passes through the origin is strong, Kramer's procedure should be adopted, otherwise use Durbin and Watson's procedure.

In Chapter 8, the DW test was found to be an approximately locally best invariant test against first-order moving average disturbances. A new exact test for first-order moving average disturbances was proposed and its small sample power properties were found to compare favourably with those of the DW test for selected design matrices, especially for comparatively large values of the moving average parameter.

Finally, in view of the encouraging power properties of this new F test against first-order moving average disturbances, we conjecture that a similar test may also perform well against first-order autoregressive disturbances.[1] It is certainly a possibility that deserves investigation.

---

1. Note that the transformation used to reduce the covariance matrix for MA(1) disturbances to a diagonal matrix, also diagonalizes a close approximation to the AR(1) covariance matrix.

ACKNOWLEDGEMENTS

The burden of writing this thesis was made bearable by the help and moral support of a number of people to whom I am most grateful.

First, I wish to acknowledge the large debt I owe to my supervisor, Professor Tony Rayner, for his encouragement and often much needed criticism of my work.

I would also like to thank Professor David Giles for kindling my interest in various problems of statistical inference associated with the linear regression model and for his kind encouragement and constructive support at all stages of the project. In addition, I am most grateful to David and his wife Lorraine for their friendly hospitality on two occasions in Melbourne.

This thesis has greatly benefited from comments and criticisms from several people. In particular, I am indebted to Dr. C. Dubbelman, Mr. Grant Hillier, Professor H. Neudecker and Dr. Tony Pakes for their detailed comments on earlier drafts of various chapters.

I am especially grateful to the University Grants Committee for financial assistance through a mixed tenure postgraduate scholarship which allowed me to spend a rewarding ten months at Monash University. In addition, I wish to thank both the Department of Economics at the University of Canterbury and the Faculty of Economics and Politics at Monash University for providing excellent office accommodation, computer time and part-time employment.

Finally, I wish to express my thanks to Jan Ottrey who cheerfully and intelligently tackled the difficult task of typing the final manuscript. The quality of the final typescript speaks volumes of her skill as a typist.

REFERENCES

Abrahamse, A.P.J. and J. Koerts (1969), "A Comparison Between the Power of the Durbin-Watson Test and the Power of the BLUS Test", *Journal of the American Statistical Association*, 64, pp. 938-948.

Abrahamse, A.P.J. and J. Koerts (1971), "New Estimators of Disturbances in Regression Analysis", *Journal of the American Statistical Association*, 66, pp. 71-74.

Abrahamse, A.P.J. and A.S. Louter (1971), "On a New Test for Autocorrelation in Least Squares Regression", *Biometrika*, 58, pp. 53-60.

Ahmad, R. (1972), "Extension of the Normal Theory to Spherical Families", *Trabajos de Estadistica y de Investigacion Operativa*, 23, pp. 51-60.

Aitchison, J. and S.D. Silvey (1960), "Maximum-likelihood Estimation Procedures and Associated Tests of Significance", *Journal of the Royal Statistical Society*, B, 22, pp. 154-171.

Almon, S. (1965), "The Distributed Lag Between Capital Appropriations and Expenditures", *Econometrica*, 33, pp. 178-196.

Anderson, E.W. and D.M. Ellis (1971), "Error Distributions in Navigation", *Journal of the Institute of Navigation*, 24, pp. 429-442.

Anderson, R.L. and T.W. Anderson (1950), "Distribution of the Circular Serial Correlation Coefficient for Residuals from a Fitted Fourier Series", *Annals of Mathematical Statistics*, 21, pp. 59-81.

Anderson, T.W. (1948), "On the Theory of Testing Serial Correlation", *Skandinavisk Aktuarietidskrift*, 31, pp. 88-116.

Anderson, T.W. and J.B. Taylor (1976), "Strong Consistency of Least Squares Estimates in Normal Linear Regression", *Annals of Statistics*, 4, pp. 788-790.

Ando, A. and F. Modigliani (1963), "The 'Life Cycle' Hypothesis of Saving: Aggregate Implications and Tests", *American Economic Review*, 53, pp. 55-84.

Andrews, D.F. (1971), "Significance Tests Based on Residuals", *Biometrika*, 58, pp. 139-148.

Anscombe, F.J. (1967), "Topics in the Investigation of Linear Relations Fitted by the Method of Least Squares", *Journal of the Royal Statistical Society*, B, 29, pp. 1-52.

Baldessari, B. (1965), "Osservazioni Sulle Funzioni Caratteristiche de Alcume Classi di V.C. n-Dimensionali", *Pubblicazioni della Seconda Serie dell'Istituto di Calcolo delle Probilità dell'Università di Roma, 49*, pp. 1-11.

Bartels, C.P.A. and L. Hordijk (1977), "On the Power of the Generalized Morgan Contiguity Coefficient in Testing for Spatial Autocorrelation Among Regression Disturbances", *Regional Science and Urban Economics*, 7, pp. 83-101.

Bartlett, M.S. (1934), "The Vector Representation of a Sample", *Proceedings of the Cambridge Philosophical Society*, 30, pp. 327-340.

Behnken, D.W. and N.R. Draper (1972), "Residuals and Their Variance Patterns", *Technometrics*, 14, pp. 101-111.

Berenblut, I.I. and G.I. Webb (1973), "A New Test for Autocorrelated Errors in the Linear Regression Model", *Journal of the Royal Statistical Society*, B, 35, pp. 33-50.

Bessel, F.W. (1818), *Fundamenta Astronomiae*. Königsberg, Nicolovius.

Bessel, F.W. (1838), "Untersuchungen über die Wahrscheinlichkeit von Beobachtungsfehlern", *Astronomische Nachrichten*, 15, pp. 369-404.

Bickel, P.J. (1978), "Using Residuals Robustly I: Tests for Heteroscedasticity, Nonlinearity", *Annals of Statistics*, 6, pp. 266-291.

Blake, I.F. and J.B. Thomas (1968), "On a Class of Processes Arising in Linear Estimation Theory", *I.E.E.E. Transactions on Information Theory*, IT-14, pp. 12-16.

Blattberg, R.C. (1973), "Evaluation of the Power of the Durbin-Watson Statistic for Non-first Order Serial Correlation Alternatives", *Review of Economics and Statistics*, 55, pp. 508-515.

Blattberg, R.C. and N.J. Gonedes (1974), "A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices", *Journal of Business*, 47, pp. 244-280.

Box, G.E.P. and J.S. Hunter (1957), "Multi-factor Experimental Designs for Exploring Response Surfaces", *Annals of Mathematical Statistics*, 28, pp. 195-241.

Box, G.E.P. and G.M. Jenkins (1970), *Time Series Analysis: Forecasting and Control*. San Francisco, Holden-Day.

Box, G.E.P. and M.E. Müller (1958), "A Note on the Generation of Random Normal Deviates", *Annals of Mathematical Statistics*, 29, pp. 610-611.

Box, G.E.P. and D.A. Pierce (1970), "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models", *Journal of the American Statistical Association*, 65, pp. 1509-1526.

Breusch, T.S. (1978), "Testing for Autocorrelation in Dynamic Linear Models", *Australian Economic Papers*, 17, pp. 334-355.

Brown, R.L., J. Durbin and J.M. Evans (1975), "Techniques for Testing the Constancy of Regression Relationships Over Time", *Journal of the Royal Statistical Society*, B, 37, pp. 149-163.

Carlson, J.A. (1975), "Are Price Expectations Normally Distributed?", *Journal of the American Statistical Association*, 70, pp. 749-754.

Chay, S.C., R.D. Fardo and M. Mazumdar (1975), "On Using the Box-Müller Transformation with Multiplicative Congruential Pseudo-random Number Generators", *Applied Statistics*, 24, pp. 132-135.

Chow, G.C. (1960), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions", *Econometrica*, 28, pp. 591-605.

Chu, K.C. (1973), "Estimation and Decision for Linear Systems with Elliptical Random Processes", *I.E.E.E. Transactions on Automatic Control*, AC-18, pp. 499-505.

Cleur, E.M. (1973), "A Simulation Study on the Powers of Three Tests for Serial Correlation in the Errors from Least Squares Analysis", *Statistica*, 33, pp. 285-300.

Cliff, A.D. and J.K. Ord (1973), *Spatial Autocorrelation*. London, Pion.

Cochrane, D. and G.H. Orcutt (1949), "Applications of Least Squares Regression to Relationships Containing Autocorrelated Error Terms", *Journal of the American Statistical Association*, 44, pp. 32-61.

Coveyou, R.R. and R.D. MacPherson (1967), "Fourier Analysis of Uniform Random Number Generators", *Journal of the Association for Computing Machinery*, 14, pp. 100-119.

Cramér, H. (1946), *Mathematical Methods of Statistics*. Princeton, Princeton University Press.

Cramér, H. and M.R. Leadbetter (1967), *Stationary and Related Stochastic Processes*. New York, John Wiley.

Crawford, J.J. (1977), "Elliptically Contoured Measures on Infinite-dimensional Banach Spaces", *Studia Mathematica*, 60, pp. 15-32.

D'Agostino, R.B. (1971), "An Omnibus Test of Normality for Moderate and Large Size Samples", *Biometrika*, 58, pp. 341-348.

Danes, M. (1975), "The Measurement and Explanation of Inflationary Expectations in Australia", *Australian Economic Papers*, 14, pp. 75-87.

Daniel, C. (1960), "Locating Outliers in Factorial Experiments", *Technometrics*, 2, pp. 149-156.

Daniel, C. and F.S. Wood (1971), *Fitting Equations to Data*. New York, Wiley-Interscience.

Das Gupta, S., M.L. Eaton, I. Olkin, M. Perlman, L.J. Savage and M. Sobel (1972), "Inequalities on the Probability Content of Convex Regions for Elliptically Contoured Distributions", *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 2, pp. 241-265.

David, H.A., H.O. Hartley and E.S. Pearson (1954), "The Distribution of the Ratio, in a Single Normal Sample, of Range to Standard Deviation", *Biometrika*, 41, pp. 482-493.

Dawid, A.P. (1977), "Spherical Matrix Distributions and a Multivariate Model", *Journal of the Royal Statistical Society*, B, 39, pp. 254-261.

Dempster, A.P. (1969), *Elements of Continuous Multivariate Analysis.* Reading, Massachusetts, Addison-Wesley.

Dent, W.T. (1973), "A Power Study of Several Tests for Autocorrelation", *New Zealand Economic Papers*, 7, pp. 109-120.

Dent, W.T. and S. Cassing (1978), "On Durbin's and Sims' Residuals in Autocorrelation Tests", *Econometrica*, 46, pp. 1489-1492.

Dent, W.T. and G.P.H. Styan (1978), "Uncorrelated Residuals from Linear Models", *Journal of Econometrics*, 7, pp. 211-225.

Drygas, H. (1971), "Consistency of the Least Squares and Gauss-Markov Estimators in Regression Models", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 17, pp. 309-326.

Drygas, H. (1976), "Weak and Strong Consistency of the Least Squares Estimators in Regression Models", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 34, pp. 119-127.

Dubbelman, C. (1972), "A Priori Fixed Covariance Matrices of Disturbance Estimators", *European Economic Review*, 3, pp. 413-436.

Dubbelman, C. (1973), "BLU Disturbance Estimators with Fixed Covariance Matrices (BLUF Disturbance Estimators)", unpublished manuscript, report 7302, Netherlands School of Economics.

Dubbelman, C. (1978), *Disturbances in the Linear Model: Estimation and Hypothesis Testing.* The Hague, Martinus Nijhoff.

Dubbelman, C., A.P.J. Abrahamse and J. Koerts (1972), "A New Class of Disturbance Estimators in the General Linear Model", *Statistica Neerlandica*, 26, pp. 127-141.

Dubbelman, C., A.S. Louter and A.P.J. Abrahamse (1978), "On Typical Characteristics of Economic Time Series and the Relative Qualities of Five Autocorrelation Tests", *Journal of Econometrics*, 8, pp. 295-306.

Durbin, J. (1960), "Estimation of Parameters in Time-series Regression Models", *Journal of the Royal Statistical Society*, B, 22, pp. 139-153.

Durbin, J. (1967), "Tests of Serial Independence Based on the Cumulated Periodogram", *Bulletin of the International Statistical Institute*, 42, pp. 1039-1048.

Durbin, J. (1969), "Tests for Serial Correlation in Regression Analysis Based on the Periodogram of Least Squares Residuals", *Biometrika*, 56, pp. 1-15.

Durbin, J. (1970a), "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors Are Lagged Dependent Variables", *Econometrica*, 38, pp. 410-421.

Durbin, J. (1970b), "An Alternative to the Bounds Test for Testing for Serial Correlation in Least Squares Regression", *Econometrica*, 38, pp. 422-429.

Durbin, J. and G.S. Watson (1950), "Testing for Serial Correlation in Least Squares Regression I", *Biometrika*, 37, pp. 409-428.

Durbin, J. and G.S. Watson (1951), "Testing for Serial Correlation in Least Squares Regression II", *Biometrika*, 38, pp. 159-178.

Durbin, J. and G.S. Watson (1971), "Testing for Serial Correlation in Least Squares Regression III", *Biometrika*, 58, pp. 1-19.

Ebbeler, D.H. (1975), "On the Probability of Correct Model Selection Using the Maximum $\bar{R}^2$ Choice Criterion", *International Economic Review*, 16, pp. 516-520.

Eddington, A.S. (1933), "Notes on the Method of Least Squares", *Proceedings of the Physical Society*, 45, pp. 271-287.

Eicker, F. (1963), "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions", *Annals of Mathematical Statistics*, 34, pp. 447-456.

Ellenberg, J.H. (1973), "The Joint Distribution of the Standardized Least Squares Residuals from a General Linear Regression", *Journal of the American Statistical Association*, 68, pp. 941-943.

Ellenberg, J.H. (1976), "Testing for Single Outlier from a General Linear Regression", *Biometrics*, 32, pp. 637-645.

Epps, T.W. and M.L. Epps (1977), "The Robustness of Some Standard Tests for Autocorrelation and Heteroskedasticity When Both Problems Are Present", *Econometrica*, 45, pp. 745-753.

Fama, E.F. (1963), "Mandelbrot and the Stable Paretian Hypothesis", *Journal of Business*, 36, pp. 420-429.

Fama, E.F. (1965), "The Behaviour of Stock-Market Prices", *Journal of Business*, 38, pp. 34-105.

Farebrother, R.W. (1978a), "Durbin-Watson Tables", unpublished manuscript, University of Manchester.

Farebrother, R.W. (1978b), "The Durbin-Watson Test for Serial Correlation When There Is No Intercept in the Regression", unpublished manuscript, University of Manchester.

Farely, J.U. and M.J. Hinich (1970), "A Test for a Shifting Slope Coefficient in a Linear Model", *Journal of the American Statistical Association*, 65, pp. 1320-1329.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications, Volume II.* 2nd ed. New York, John Wiley.

Ferguson, T.S. (1961), "On the Rejection of Outliers", *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1, pp. 253-287.

Ferguson, T.S. (1967), *Mathematical Statistics: A Decision Theoretic Approach*. New York, Academic Press.

Fisher, F.M. (1970), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note", *Econometrica*, 38, pp. 361-366.

Fisz, M. (1963), *Probability Theory and Mathematical Statistics*. 3rd ed. New York, John Wiley.

Fitts, J. (1973), "Testing for Autocorrelation in the Autoregressive Moving Average Error Model", *Journal of Econometrics*, 1, pp. 363-376.

Fraser, D.A.S., I. Guttman and G.P.H. Styan (1976), "Serial Correlation and Distributions on the Sphere", *Communications in Statistics*, A, A5, pp. 97-118.

Gauss, C.F. (1809), *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium.* Hamburg, Perthes and Besser. [Reprinted 1906 in *Werke*, 7. Göttingen, Königlichen Gesellschaft der Wissenschaften, pp. 1-288.]

Geary, R.C. (1947), "Testing for Normality", *Biometrika*, 34, pp. 209-42.

Geary, R.C. (1970), "Relative Efficiency of Count of Sign Changes for Assessing Residual Autoregression in Least Squares Regression", *Biometrika*, 57, pp. 123-127.

Ghosh, M. and E. Pollak (1975), "Some Properties of Multivariate Distributions with P.D.F.'s Constant on Ellipsoids", *Communications in Statistics*, 4, pp. 1157-1160.

Glejser, H. (1969), "A New Test for Heteroscedasticity", *Journal of the American Statistical Association*, 64, pp. 316-323.

Godfrey, L.G. (1973), "A Note on the Treatment of Serial Correlation", *Canadian Journal of Economics*, 6, pp. 567-573.

Godfrey, L.G. (1974), "A Further Note on the Treatment of Serial Correlation", *Canadian Journal of Economics*, 7, pp. 673-676.

Godfrey, L.G. (1978a), "A Note on the Use of Durbin's h Test When the Equation Is Estimated by Instrumental Variables", *Econometrica*, 46, pp. 225-228.

Godfrey, L.G. (1978b), "Testing for Multiplicative Heteroskedasticity", *Journal of Econometrics*, 8, pp. 227-236.

Godfrey, L.G. (1978c), "Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables", *Econometrica*, 46, pp. 1293-1301.

Godfrey, L.G. (1978d), "Testing for Higher Order Serial Correlation in Regression Equations When the Regressors Include Lagged Dependent Variables", *Econometrica*, 46, pp. 1303-1310.

Godfrey, L.G. and A.R. Tremayne (1978), "Testing for Fourth Order Autocorrelation in Dynamic Quarterly Regression Equations: Asymptotic Tests and Monte Carlo Evidence", unpublished manuscript, University of York.

Golder, E.R. (1976a), "Algorithm AS98: The Spectral Test for the Evaluation of Congruential Pseudo-random Generators", *Applied Statistics*, 25, pp. 173-180.

Golder, E.R. (1976b), "Remark ASR18: The Spectral Test for the Evaluation of Congruential Pseudo-random Generators", *Applied Statistics*, 25, p. 324.

Golder, E.R. and J.G. Settle (1976), "The Box-Müller Method for Generating Pseudo-random Normal Deviates", *Applied Statistics*, 25, pp. 12-20.

Goldfeld, S.M. and R.E. Quandt (1965), "Some Tests for Homoscedasticity", *Journal of the American Statistical Association*, 60, pp. 539-547.

Goldfeld, S.M. and R.E. Quandt (1972), *Nonlinear Methods in Econometrics*. Amsterdam, North-Holland.

Goldman, J. (1972), "Moments of the Sum of Circularly Symmetric Random Variables", *I.E.E.E. Transactions on Information Theory*, IT-18, pp. 297-298.

Goldman, J. (1976), "Detection in the Presence of Spherically Symmetric Random Vectors", *I.E.E.E. Transactions on Information Theory*, IT-22, pp. 52-59.

Goldsmith, P.L. and R. Boddy (1973), "Critical Analysis of Factorial Experiments and Orthogonal Fractions", *Applied Statistics*, 22, pp. 141-160.

Golub, G.H. (1965), "Numerical Methods for Solving Linear Least Squares Problems", *Numerische Mathematik*, 7, pp. 206-216.

Golub, G.H. and G.P.H. Styan (1973), "Numerical Computations for Univariate Linear Models", *Journal of Statistical Computation and Simulation*, 2, pp. 253-274.

Granger, C.W.J. (1966), "The Typical Spectral Space of an Economic Variable", *Econometrica*, 34, pp. 150-161.

Granger, C.W.J. and D. Orr (1972), "'Infinite Variance' and Research Strategy in Time Series Analysis", *Journal of the American Statistical Association*, 67, pp. 275-285.

Graybill, F.A. (1969), *Introduction to Matrices with Applications in Statistics*. Belmont, California, Wadsworth.

Griliches, Z., G.S. Maddala, R. Lucas and N. Wallace (1962), "Notes on Estimated Aggregate Quarterly Consumption Functions", *Econometrica*, 30, pp. 491-500.

Gualtierotti, A.F. (1974), "Some Remarks on Spherically Invariant Distributions", *Journal of Multivariate Analysis*, 4, pp. 347-349.

Gualtierotti, A.F. (1976), "A Likelihood Ratio Formula for Spherically Invariant Processes", *I.E.E.E. Transactions on Information Theory*, IT-22, p. 610.

Haavelmo, T. (1944), "The Probability Approach in Econometrics", *Econometrica*, 12, Supplement.

Habibagahi, H. and J.L. Pratschke (1972), "A Comparison of the Power of the von Neumann Ratio, Durbin-Watson and Geary Tests", *Review of Economics and Statistics*, 54, pp. 179-185.

Hagen, G.H.L. (1837), *Grundzüge der Wahrscheinlichkeitsrechnung*. Berlin, Ernst and Korn.

Hall, V.B. and M.L. King (1976), "Inflationary Expectations in New Zealand: A Preliminary Study", *New Zealand Economic Papers*, 10, pp. 118-151.

Hampel, F.R. (1973), "Robust Estimation: A Condensed Partial Survey", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 27, pp. 87-104.

Hannan, E.J. (1955), "Exact Tests for Serial Correlation", *Biometrika*, 42, pp. 133-142.

Hannan, E.J. (1957), "Testing for Serial Correlation in Least Squares Regression", *Biometrika*, 44, pp. 57-66.

Hannan, E.J. and R.D. Terrell (1968), "Testing for Serial Correlation After Least Squares Regression", *Econometrica*, 36, pp. 133-150.

Harrison, M.J. (1972), "On Testing for Serial Correlation in Regression When the Bounds Test Is Inconclusive", *Economic and Social Review*, 4, pp. 41-57.

Harrison, M.J. (1975), "The Power of the Durbin-Watson and Geary Tests: Comment and Further Evidence", *Review of Economics and Statistics*, 57, pp. 377-379.

Harrison, M.J. and B.P.M. McCabe (1975), "Autocorrelation with Heteroscedasticity: A Note on the Robustness of the Durbin-Watson, Geary and Henshaw Tests", *Biometrika*, 62, pp. 214-215.

Hart, B.I. (1942), "Tabulation of Probabilities for the Ratio of the Mean Square Successive Difference to the Variance", *Annals of Mathematical Statistics*, 13, pp. 207-214.

Harter, H.L. (1974a), "The Method of Least Squares and Some Alternatives - Part I", *International Statistical Review*, 42, pp. 147-174.

Harter, H.L. (1974b), "The Method of Least Squares and Some Alternatives - Part II", *International Statistical Review*, 42, pp. 235-264.

Harter, H.L. (1975a), "The Method of Least Squares and Some Alternatives - Part III", *International Statistical Review*, 43, pp. 1-44.

Harter, H.L. (1975b), "The Method of Least Squares and Some Alternatives - Part IV", *International Statistical Review*, 43, pp. 125-190.

Harter, H.L. (1975c), "The Method of Least Squares and Some Alternatives - Part V", *International Statistical Review*, 43, pp. 269-278.

Harvey, A.C. (1976), "Estimating Regression Models with Multiplicative Heteroscedasticity", *Econometrica*, 44, pp. 461-466.

Harvey, A.C. and P. Collier (1977), "Testing for Functional Misspecification in Regression Analysis", *Journal of Econometrics*, 6, pp. 103-119.

Harvey, A.C. and G.D.A. Phillips (1974), "A Comparison of the Power of Some Tests for Heteroskedasticity in the General Linear Model", *Journal of Econometrics*, 2, pp. 307-316.

Heyadat, A. and D.S. Robson (1970), "Independent Stepwise Residuals for Testing Homoscedasticity", *Journal of the American Statistical Association*, 65, pp. 1573-1581.

Hoaglin, D.C. (1976), "Theoretical Properties of Congruential Random-Number Generators: An Empirical View", Presented at the ORSA/TIMS Joint National Meeting, Miami, Florida.

Hoaglin, D.C. and M.L. King (1978), "Remark ASR24: A Remark on Algorithm AS98: The Spectral Test for the Evaluation of Congruential Pseudo-random Generators", *Applied Statistics*, 27, pp. 375-377.

Hsu, D.A. (1979), "Long-tailed Distributions for Position Errors in Navigation", *Applied Statistics*, 28, pp. 62-72.

Huang, C.J. (1973), "Normality of Disturbance Terms and Specification of the Production Functional Forms", *Southern Economic Journal*, 40, pp. 12-18.

Huang, C.J. and B.W. Bolch (1974), "On the Testing of Regression Disturbances for Normality", *Journal of the American Statistical Association*, 69, pp. 330-335.

Huber, P.J. (1972), "Robust Statistics: A Review", *Annals of Mathematical Statistics*, 43, pp. 1041-1067.

Huber, P.J. (1973), "Robust Regression: Asymptotics, Conjectures and Monte Carlo", *Annals of Statistics*, 1, pp. 799-821.

Imhof, P.J. (1961), "Computing the Distribution of Quadratic Forms in Normal Variables", *Biometrika*, 48, pp. 419-426.

Jayatissa, W.A. (1977), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions When Disturbance Variances are Unequal", *Econometrica*, 45, pp. 1291-1292.

Jeffreys, H. (1938), "The Law of Error and the Combination of Observations", *Philosophical Transactions of the Royal Society of London*, A, 237, pp. 231-271.

Jeffreys, H. (1948), *Theory of Probability*. 2nd ed. London, Oxford University Press.

Jennrich, R.I. (1969), "Asymptotic Properties of Non-linear Least Squares Estimators", *Annals of Mathematical Statistics*, 40, pp. 633-643.

Jensen, D.R. (1978), "On Consistent Cauchy Averages", *Journal of Statistical Computation and Simulation*, 7, pp. 287-289.

Johnston, J. (1972), *Econometric Methods*. 2nd ed. New York, McGraw-Hill.

Kadiyala, K.R. (1970), "Testing for the Independence of Regression Disturbances", *Econometrica*, 38, pp. 97-117.

Kariya, T. (1977), "A Robustness Property of the Tests for Serial Correlation", *Annals of Statistics*, 5, pp. 1212-1220.

Kariya, T. and M.L. Eaton (1977), "Robust Tests for Spherical Symmetry", *Annals of Statistics*, 5, pp. 206-215.

Kelker, D. (1970), "Distribution Theory of Spherical Distributions and a Location-scale Parameter Generalization", *Sankhyā*, A, 32, pp. 419-430.

Kenkel, J.L. (1974), "Some Small-Sample Properties of Durbin's Tests for Serial Correlation in Regression Models Containing Lagged Dependent Variables", *Econometrica*, 42, pp. 763-769.

Kenkel, J.L. (1975), "Small-Sample Tests for Serial Correlation in Models Containing Lagged Dependent Variables", *Review of Economics and Statistics*, 57, pp. 383-386.

Kenkel, J.L. (1976), "Comment on the Small-Sample Power of Durbin's h-Test", *Journal of the American Statistical Association*, 71, pp. 96-97.

Kenward, L.R. (1975), "Autocorrelation and Dynamic Methodology with an Application to Wage Determination Models", *Journal of Econometrics*, 3, pp. 179-188.

Kenward, L.R. (1976), "A New Test for Residual Randomness in a Class of Dynamic Autocorrelated Econometric Models", *Canadian Journal of Statistics*, 4, pp. 51-64.

King, M.L. and D.E.A. Giles (1977), "A Note on Wallis' Bounds Test and Negative Autocorrelation", *Econometrica*, 45, pp. 1023-1026.

King, M.L. and D.E.A. Giles (1978), "A Comparison of Some Tests for Fourth-order Autocorrelation", *Australian Economic Papers*, 17, pp. 323-333.

Kingman, J.F.C. (1963), "Random Walks with Spherical Symmetry", *Acta Mathematica*, 109, pp. 11-53.

Kingman, J.F.C. (1972), "On Random Sequences with Spherical Symmetry", *Biometrika*, 59, pp. 492-494.

Knöbl, A. (1974), "Price Expectations and Actual Price Behaviour in Germany", *International Monetary Fund Staff Papers*, 21, pp. 83-100.

Knopp, K. (1956), *Infinite Sequences and Series*. New York, Dover.

Knuth, D.E. (1969), *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*. Reading, Massachusetts, Addison-Wesley.

Koenker, R. and G. Bassett (1978), "Regression Quantiles", *Econometrica*, 46, pp. 33-50.

Koerts, J. and A.P.J. Abrahamse (1968), "On the Power of the BLUS Procedure", *Journal of the American Statistical Association*, 63, pp. 1227-1236.

Koerts, J. and A.P.J. Abrahamse (1969), *On the Theory and Application of the General Linear Model*. Rotterdam, Rotterdam University Press.

Kramer, G. (1971), "On the Durbin-Watson Bounds Test in the Case of Regression Through the Origin", *Jahrbucher für Nationalökonomie und Satistik*, 185, pp. 345-358.

Lai, T.L. and H. Robbins (1977), "Strong Consistency of Least-Squares Estimates in Regression Models", *Proceedings of the National Academy of Science*, 74, pp. 2667-2669.

Lai, T.L., H. Robbins and C.Z. Wei (1978), "Strong Consistency of Least Squares Estimates in Multiple Regression", *Proceedings of the National Academy of Science*, 75, pp. 3034-3036.

Laplace, P.S. (1812), *Théorie Analytique des Probabilités*. Paris, Courcier.

Lehmann, E.L. (1959), *Testing Statistical Hypotheses*. New York, John Wiley.

Lehmann, E.L. and C. Stein (1948), "Most Powerful Tests of Composite Hypotheses, I. Normal Distributions", *Annals of Mathematical Statistics*, 19, pp. 495-516.

L'Esperance, W.L. and D. Taylor (1975), "The Power of Four Tests of Autocorrelation in the Linear Regression Model", *Journal of Econometrics*, 3, pp. 1-21.

Lindeberg, J.W. (1922), "Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung", *Mathematische Zeitschrift*, 15, pp. 211-225.

Lindley, D.V. (1979), "Discussion of Professor Dawid's Paper: 'Conditional Independence in Statistical Theory'", *Journal of the Royal Statistical Society*, B, 41, pp. 15-16.

Lord, R.D. (1954), "The Use of the Hankel Transform in Statistics I. General Theory and Examples", *Biometrika*, 41, pp. 44-55.

Louter, A.S. and C. Dubbelman (1973), "An Exact Autocorrelation Test for Small n and k, a Computer Program and a Table of Significance Points", unpublished manuscript, report 7304, Netherlands School of Economics.

Louter, A.S. and J. Koerts (1970), "On the Kuiper Test for Normality with Mean and Variance Unknown", *Statistica Neerlandica*, 24, pp. 83-87.

Lund, R.E. (1975), "Tables for an Approximate Test for Outliers in Linear Models", *Technometrics*, 17, pp. 473-476.

Lyons, N.I. and C.H. Proctor (1977), "A Test for Regression Function Adequacy", *Communications in Statistics*, A, A6, pp. 81-86.

Maddala, G.S. and A.S. Rao (1973), "Tests for Serial Correlation in Regression Models with Lagged Dependent Variables and Serially Correlated Errors", *Econometrica*, 41, pp. 761-774.

Magnus, J.R. (1978), "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix", *Journal of Econometrics*, 7, pp. 281-312.

Malinvaud, E. (1970), *Statistical Methods of Econometrics*. 2nd ed. Amsterdam, North-Holland.

Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices", *Journal of Business*, 36, pp. 394-419.

Mandelbrot, B. (1967), "The Variation of Other Speculative Prices", *Journal of Business*, 40, pp. 393-413.

Mann, H.B. and A. Wald (1942), "On the Choice of the Number of Class Intervals in the Application of the Chi Square Test", *Annals of Mathematical Statistics*, 13, pp. 306-317.

Maxwell, J.C. (1860), "Illustrations of the Dynamical Theory of Gases. Part 1: On the Motions and Collisions of Perfectly Elastic Spheres", *Philosophical Magazine*, 19, pp. 19-32.

McGraw, D.K. and J.F. Wagner (1968), "Elliptically Symmetric Distributions", *I.E.E.E. Transactions on Information Theory*, IT-14, pp. 110-120.

Mehta, M.L. (1967), *Random Matrices and the Statistical Theory of Energy Levels*. New York, Academic Press.

Merrington, M. and C.M. Thompson (1943), "Tables of Percentage Points of the Inverted Beta (F) Distribution", with prefatory note by E.S. Pearson, *Biometrika*, 33, pp. 73-88.

Miller, K.S. (1964), *Multidimensional Gaussian Distributions*. New York, John Wiley.

Mukantseva, L.A. (1977), "Testing Normality in One-Dimensional and Multidimensional Linear Regression", *Theory of Probability and its Applications*, 22, pp. 591-602.

Neave, H.R. (1973), "On Using the Box-Müller Transformation with Multiplicative Congruential Pseudo-random Number Generators", *Applied Statistics*, 22, pp. 92-97.

Nerlove, M. and K.F. Wallis (1966), "Use of the Durbin-Watson Statistic in Inappropriate Situations", *Econometrica*, 34, pp. 235-238.

Neudecker, H. (1977), "Abrahamse and Koerts' 'New Estimator' of Disturbances in Regression Analysis", *Journal of Econometrics*, 5, pp. 129-133.

Newcomb, S. (1886), "A Generalized Theory of the Combination of Observations so as to Obtain the Best Result", *American Journal of Mathematics*, 8, pp. 343-366.

Nicholls, D.F., A.R. Pagan and R.D. Terrell (1975), "The Estimation and Use of Models with Moving Average Disturbance Terms: A Survey", *International Economic Review*, 16, pp. 113-134.

Park, S. (1972), "Some Small-Sample Properties of Durbin's h Statistic", *American Statistical Association 1972 Proceedings of the Business and Economic Statistics Section*, pp. 413-419.

Park, S. (1975), "On the Small-Sample Power of Durbin's h Test", *Journal of the American Statistical Association*, 70, pp. 60-63.

Park, S. (1976), "Rejoinder to 'Comment on the Small-Sample Power of Durbin's h-Test'", *Journal of the American Statistical Association*, 71, pp. 97-98.

Parzen, E. (1960), *Modern Probability Theory and Its Applications*. New York, John Wiley.

Pearson, E.S. and H.O. Hartley (1954), *Biometrika Tables for Statisticians, Volume 1*. Cambridge, Biometrika Trust.

Pesaran, M.H. (1973), "Exact Maximum Likelihood Estimation of a Regression Equation with a First-Order Moving-Average Error", *Review of Economic Studies*, 40, pp. 529-535.

Phillips, G.D.A. and A.C. Harvey (1974), "A Simple Test for Serial Correlation in Regression Analysis", *Journal of the American Statistical Association*, 69, pp. 935-939.

Picinbono, B. (1970), "Spherically Invariant and Compound Gaussian Stochastic Processes", *I.E.E.E. Transactions on Information Theory*, IT-16, pp. 77-79.

Poincaré, H. (1912), *Calcul des Probabilités*. 2nd ed. Paris, Gauthier-Villars.

Poisson, S.D. (1824), "Sur la Probabilité des Résultats Moyens des Observations", *Connaissance des Tems*, Année 1827, pp. 273-302.

Praetz, P.D. (1972), "The Distribution of Share Price Changes", *Journal of Business*, 45, pp. 49-55.

Praetz, P.D. and E.J.G. Wilson (1978), "The Distribution of Stock Market Returns: 1958-1973", *Australian Journal of Management*, 3, pp. 79-90.

Prescott, P. (1975), "An Approximate Test for Outliers in Linear Models", *Technometrics*, 17, pp. 129-132.

Press, S.J. (1967), "A Compound Events Model for Security Prices", *Journal of Business*, 40, pp. 317-335.

Putter, J. (1967), "Orthonormal Bases of Error Spaces and Their Use for Investigating the Normality and Variances of Residuals", *Journal of the American Statistical Association*, 62, pp. 1022-1036.

Ramsey, J.B. (1969), "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis", *Journal of the Royal Statistical Society*, B, 31, pp. 350-371.

Ramsey, J.B. and R. Gilbert (1972), "A Monte Carlo Study of Some Small-Sample Properties of Tests for Specification Error", *Journal of the American Statistical Association*, 67, pp. 180-186.

Ramsey, J.B. and P. Schmidt (1976), "Some Further Results on the Use of OLS and BLUS Residuals in Specification Error Tests", *Journal of the American Statistical Association*, 71, pp. 389-390.

Ramsey, J.B. and P. Zarembka (1971), "Specification Error Tests and Alternative Functional Forms of the Aggregate Production Function", *Journal of the American Statistical Association*, 66, pp. 471-477.

Rao, C.R. (1973), *Linear Statistical Inference and Its Applications*. 2nd ed. New York, John Wiley.

Rowley, J.C.R. and D.A. Wilton (1973), "Quarterly Models of Wage Determination: Some New Efficient Estimates", *American Economic Review*, 63, pp. 380-389.

Sargan, J.D. (1959), "The Estimation of Relationships with Autocorrelated Residuals by the Use of Instrumental Variables", *Journal of the Royal Statistical Society*, B, 21, pp. 91-105.

Särndal, C. (1971), "Studies in the History of Probability and Statistics XXVII. The Hypothesis of Elementary Errors and the Scandinavian School in Statistical Theory", *Biometrika*, 58, pp. 375-391.

Savin, N.E. and K.J. White (1977), "The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors", *Econometrica*, 45, pp. 1989-1996.

Savin, N.E. and K.J. White (1978), "Testing for Autocorrelation with Missing Observations", *Econometrica*, 46, pp. 59-67.

Scheffé, H. (1959), *The Analysis of Variance*. New York, John Wiley.

Schmidt, P. (1972), "A Generalization of the Durbin-Watson Test", *Australian Economic Papers*, 11, pp. 203-209.

Schmidt, P. (1976), *Econometrics*. New York, Marcel Dekker.

Schmidt, P. and D.K. Guilkey (1975), "Some Further Evidence on the Power of the Durbin-Watson and Geary Tests", *Review of Economics and Statistics*, 57, pp. 379-382.

Schmidt, P. and R. Sickles (1977), "Some Further Evidence on the Use of the Chow Test under Heteroskedasticity", *Econometrica*, 45, pp. 1293-1298.

Seber, G.A.F. (1977), *Linear Regression Analysis*. New York, John Wiley.

Shapiro, S.S. and M.B. Wilk (1965), "An Analysis of Variance Test for Normality (Complete Samples)", *Biometrika*, 52, pp. 591-611.

Shapiro, S.S., M.B. Wilk and H.J. Chen (1968), "A Comparative Study of Various Tests for Normality", *Journal of the American Statistical Association*, 63, pp. 1343-1372.

Silvey, S.D. (1959), "The Lagrangian Multiplier Test", *Annals of Mathematical Statistics*, 30, pp. 389-407.

Sims, C.A. (1974), "Distributed Lags", in M.D. Intriligator and D.A. Kendrick (ed.). *Frontiers of Quantitative Economics, Volume II*. Amsterdam, North-Holland, pp. 289-338.

Sims, C.A. (1975), "A Note on Exact Tests for Serial Correlation", *Journal of the American Statistical Association*, 70, pp. 162-165.

Smith, V.K. (1976), "The Estimated Power of Several Tests for Autocorrelation with Non-first-order Alternatives", *Journal of the American Statistical Association*, 71, pp. 879-883.

Smith, V.K. (1977a), "A Note on the Power of the Cumulated Periodogram Test for Autocorrelation", *European Economic Review*, 9, pp. 373-377.

Smith, V.K. (1977b), "Further Evidence on the Power of Conventional Tests for Autocorrelation", unpublished manuscript, Resources for the Future, Washington.

Spencer, B.G. (1975), "The Small Sample Bias of Durbin's Tests for Serial Correlation When One of the Regressors Is the Lagged Dependent Variable and the Null Hypothesis Is True", *Journal of Econometrics*, 3, pp. 249-254.

Stefansky, W. (1971), "Rejecting Outliers by Maximum Normed Residual", *Annals of Mathematical Statistics*, 42, pp. 35-45.

Stefansky, W. (1972), "Rejecting Outliers in Factorial Designs", *Technometrics*, 14, pp. 469-479.

Stephens, M.A. (1970), "Use of the Kolmogorov-Smirnov, Cramér-von Mises and Related Statistics Without Extensive Tables", *Journal of the Royal Statistical Society*, B, 32, pp. 115-122.

Stout, W.F. (1974), *Almost Sure Convergence*. New York, Academic Press.

Strawderman, W.E. (1974), "Minimax Estimation of Location Parameters for Certain Spherically Symmetric Distributions", *Journal of Multivariate Analysis*, 4, pp. 255-264.

Swed, F.S. and C. Eisenhart (1943), "Tables for Testing Randomness of Grouping in a Sequence of Alternatives", *Annals of Mathematical Statistics*, 14, pp. 66-87.

Swick, D.A. (1974), Letter to the Editors, *Applied Statistics*, 23, p. 233.

Taylor, L.D. and T.A. Wilson (1964), "Three-pass Least Squares: A Method for Estimating Models with a Lagged Dependent Variable", *Review of Economics and Statistics*, 46, pp. 329-346.

Theil, H. (1965), "The Analysis of Disturbances in Regression Analysis", *Journal of the American Statistical Association*, 60, pp. 1067-1079.

Theil, H. (1968), "A Simplification of the BLUS Procedure for Analyzing Regression Disturbances", *Journal of the American Statistical Association*, 63, pp. 242-251.

Theil, H. (1971), *Principles of Econometrics*. New York, John Wiley.

Theil, H. and A.L. Nagar (1961), "Testing the Independence of Regression Disturbances", *Journal of the American Statistical Association*, 56, pp. 793-806.

Thomas, J.J. and K.F. Wallis (1971), "Seasonal Variation in Regression Analysis", *Journal of the Royal Statistical Society*, A, 134, pp. 57-72.

Tiao, G.C. and I. Guttman (1967), "Analysis of Outliers with Adjusted Residuals", *Technometrics*, 9, pp. 541-559.

Tietjen, G.L., R.H. Moore and R.J. Beckman (1973), "Testing for a Single Outlier in Simple Linear Regression", *Technometrics*, 15, pp. 717-721.

Tillman, J.A. (1975), "The Power of the Durbin-Watson Test", *Econometrica*, 43, pp. 959-974.

Toyoda, T. (1974), "Use of the Chow Test under Heteroscedasticity", *Econometrica*, 42, pp. 601-608.

Turnovsky, S.J. and M.L. Wachter (1972), "A Test of the 'Expectations Hypothesis' Using Directly Observed Wage and Price Expectations", *Review of Economics and Statistics*, 54, pp. 47-54.

Vershik, A.M. (1964),"Some Characteristic Properties of Gaussian Stochastic Processes", *Theory of Probability and its Applications*, 9, pp. 353-356.

Vinod, H.D. (1973), "Generalization of the Durbin-Watson Statistic for Higher Order Autoregressive Processes", *Communications in Statistics*, 2, pp. 115-144.

von Neumann, J. (1941), "Distribution of the Ratio of the Mean Square Successive Difference to the Variance", *Annals of Mathematical Statistics*, 22, pp. 367-395.

Wald, A. and J. Wolfowitz (1940), "On a Test Whether Two Samples Are from the Same Population", *Annals of Mathematical Statistics*, 11, pp. 147-162.

Wallace, T.D. (1977), "Pretest Estimation in Regression: A Survey", *American Journal of Agricultural Economics*, 59, pp. 431-443.

Wallis, K.F. (1972), "Testing for Fourth Order Autocorrelation in Quarterly Regression Equations", *Econometrica*, 40, pp. 617-636.

Ward, L.L. (1973), "Is Uncorrelating the Residuals Worth It?", unpublished M.Sc. thesis, McGill University, Quebec.

Waugh, F.V. and M.E. Abel, (1967), "On Fractional Powers of a Matrix", *Journal of the American Statistical Association*, 62, pp. 1018-1021.

Webb, G.I. (1973), "Autocorrelations and the General Linear Model", unpublished Ph.D. thesis, The City University, London.

White, A.J. (1968), *Real Analysis: An Introduction.* London, Addison-Wesley.

Williams, D.A. (1973), Letter to the Editors, *Applied Statistics*, 22, pp. 407-408.

Wise, G.L. and N.C. Gallagher (1978), "On Spherically Invariant Random Processes", *I.E.E.E. Transactions on Information Theory*, IT-24, pp. 118-120.

Wolfe, S.J. (1975), "On the Unimodality of Spherically Symmetric Stable Distribution Functions", *Journal of Multivariate Analysis*, 5, pp. 236-242.

Yao, K. (1973), "A Representation Theorem and Its Applications to Spherically-Invariant Random Processes", *I.E.E.E. Transactions on Information Theory*, IT-19, pp. 600-608.

Zellner, A. (1976), "Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student-t Error Terms", *Journal of the American Statistical Association*, 71, pp. 400-405.

Zellner, A. and C. Montmarquette (1971), "A Study of Some Aspects of Temporal Aggregation Problems in Econometric Analyses", *Review of Economics and Statistics*, 53, pp. 335-342.

APPENDIX 1

ARTIFICIALLY GENERATED TIME SERIES

DATA USED IN CHAPTERS 7 AND 8

The symbols used here, and the method of generation of the data are described in Section 3 of Chapter 7.

(i) Stationary Autoregressive Time Series Data

t	x_{1t}	x_{2t}	t	x_{1t}	x_{2t}
1	0.86893	-0.92189	31	-0.62885	-0.98272
2	1.27628	1.05377	32	-0.46272	-0.47822
3	0.36353	0.79509	33	-0.24589	1.49891
4	0.31145	0.11437	34	0.53389	0.80558
5	0.49335	1.14822	35	-0.46462	0.40587
6	-0.62251	0.10722	36	0.74724	0.53986
7	-0.70172	0.84068	37	-0.74342	0.15622
8	0.03861	-0.40973	38	-1.76168	-1.13941
9	0.23036	-1.97409	39	-1.50076	0.36781
10	0.08789	-0.36209	40	-1.52810	0.39845
11	2.64482	-1.79864	41	-1.49334	0.10871
12	-0.30889	-1.41011	42	0.98632	0.90448
13	0.82215	-0.17282	43	0.27402	1.74347
14	1.20652	1.54171	44	1.89378	3.12688
15	1.88242	1.33399	45	1.63795	1.27491
16	-0.29356	-0.79952	46	0.99448	1.25265
17	-0.99069	0.77016	47	1.81697	1.76013
18	-0.85466	0.90615	48	1.59450	1.61954
19	0.98890	-0.84103	49	1.19289	1.12912
20	-0.33641	-0.65136	50	-0.74034	-0.48832
21	-2.01937	-1.89396	51	0.35837	-1.11418
22	-2.94319	0.46659	52	-0.30379	-1.45913
23	-0.32117	-0.82075	53	0.89173	-0.60500
24	0.45328	-0.01608	54	0.55841	0.42332
25	0.34613	-1.46706	55	1.38352	-0.55158
26	0.02994	0.05691	56	0.96820	0.19215
27	0.22566	0.56146	57	1.92519	0.47906
28	-1.04088	-1.88189	58	2.88618	-1.07531
29	-0.98287	-1.97014	59	1.69062	-0.81571
30	-1.65096	-1.34578	60	1.14495	-1.67034

(ii) Linear Trend with Autoregressive Disturbances Time Series Data

t	x_{1t}	x_{2t}	t	x_{1t}	x_{2t}
1	1.11893	-0.67189	31	7.12115	6.76728
2	1.77628	1.55377	32	7.53728	7.52178
3	1.11353	1.54509	33	8.00411	9.74891
4	1.31145	1.11437	34	9.03389	9.30558
5	1.74335	2.39822	35	8.28538	9.15587
6	0.87749	1.60722	36	9.74724	9.53986
7	1.04828	2.59068	37	8.50658	9.40622
8	2.03861	1.59027	38	7.73832	8.36059
9	2.48036	0.27591	39	8.24924	10.11781
10	2.58789	2.13791	40	8.47190	10.39845
11	5.39482	0.95136	41	8.75666	10.35871
12	2.69111	1.58989	42	11.48632	11.40448
13	4.07215	3.07718	43	11.02402	12.49347
14	4.70652	5.04171	44	12.89378	14.12688
15	5.63242	5.08399	45	12.88795	12.52491
16	3.70644	3.20048	46	12.49448	12.75265
17	3.25931	5.02016	47	13.56697	13.51013
18	3.64534	5.40615	48	13.59450	13.61954
19	5.73890	3.90897	49	13.44289	13.37912
20	4.66359	4.34864	50	11.75966	12.01168
21	3.23063	3.35604	51	13.10837	11.63582
22	2.55681	5.96659	52	12.69621	11.54087
23	5.42883	4.92925	53	14.14173	12.64500
24	6.45328	5.98392	54	14.05841	13.92332
25	6.59613	4.78294	55	15.13352	13.19842
26	6.52994	6.55691	56	14.96820	14.19215
27	6.97566	7.31146	57	16.17519	14.72906
28	5.95912	5.11811	58	17.38618	13.42469
29	6.26713	5.27986	59	16.44062	13.93429
30	5.84904	6.14522	60	16.14495	13.32966

APPENDIX 2

AN EVALUATION OF THE NORMAL PSEUDO-RANDOM NUMBER GENERATOR USED IN THIS THESIS

1. INTRODUCTION

Box and Müller (1958) proved that if v_1 and v_2 are independent uniform random variables with range $(0,1)$, then z_1 and z_2 defined by

$$(A2.1) \qquad \begin{aligned} z_1 &= (-2\ell n(v_1))^{\frac{1}{2}}\sin(2\pi v_2) \\ z_2 &= (-2\ell n(v_1))^{\frac{1}{2}}\cos(2\pi v_2), \end{aligned}$$

are independent, standard normal, random variables. Standard normal pseudo-random variates used in this thesis were generated by applying this transformation to uniform pseudo-random variates with range $[0,1)$ produced by RANDOM - an intrinsic subroutine which is part of the software of the Burroughs B6700 computer. RANDOM generates uniform pseudo-random variates $(u_i, i=1,2,...)$ by the mixed congruential method:

$$|x_{n+1}| = a|x_n| + c \qquad \pmod{m}$$

$$u_n = |x_n|/m ,$$

where x_{n+1} has the same sign as x_n. x_0 is known as the seed of the integer sequence $\{x_i, i=1,...\}$. The actual values taken by the integer constants a, c and m are

$$m = 2^{39}$$

$$c = 116177073375$$

$$a = \begin{cases} 152587890725 & \text{if } x_0 \geqslant 0 , \\ 277626315293 & \text{if } x_0 < 0 . \end{cases}$$

The two values of a means that two different uniform pseudo-random number generators are available, the choice of generator being made by the sign of the seed.

Note that RANDOM is capable of generating zero values. This is unfortunate because if such a value is taken by v_1, the computer will refuse to calculate a value for $\ln(v_1)$ causing the Box-Müller transformation to break down. In order to overcome this problem, one could either test for zero v_1 values before applying (A2.1) or take the attitude that if a breakdown does occur, the experiment will be rerun using a different seed. As both uniform generators are full period generators, the relative frequency of a breakdown is extremely small - namely 2^{-40}. For this reason we adopted the latter approach and fortunately did not experience a breakdown.

Recently, Neave (1973) conjectured that the use of the Box-Müller transformation in conjunction with a congruential generator is often inadvisable since it may result in unsatisfactory sampling distributions - especially in the tails. Aspects of his analysis and conclusions have been challenged by Swick (1974), Chay et al. (1975) and Golder and Settle (1976), though there is general agreement that Neave's results provide a valid warning against the blind use of such generators. There is no doubt that bad congruential generators which produce unsatisfactory normal pseudo-random variates after the application of the Box-Müller transformation do exist.[1] It appears the Box-Müller transformation tends to magnify any deficiencies of its attendant congruential generator. There is no evidence to suggest that a congruential generator which produces satisfactory uniform pseudo-random deviates will not generate adequate standard normal deviates when coupled with the Box-Müller transformation.

We tested the two mixed congruential generators by applying Coveyou and MacPherson's (1967) spectral test to each of them. This was achieved using Golder's (1976a, 1976b) SPEC subroutine modified as suggested by Hoaglin and King (1978). The spectral test is a theoretical test which is applied to the whole cycle of a particular generator and to quote Knuth (1969, p.82), "is especially significant because not only do all good random number generators pass it, but also all linear congruential sequences now known to be bad actually fail it!" In these respects, it differs from empirical tests which, until recently,[2] were the sole means used for evaluating the performance of congruential generators but which occasionally have been found to have low power against generators with certain types of defects.

As well as testing the quality of the two uniform generators, we applied a number of empirical tests to the normal pseudo-random variates that result from applying the Box-Müller transformation in each of the four following ways:

---

1. For example, generators with small values of a should be avoided.

2. Roughly within the last ten years. For further details see Hoaglin (1976).

(i) v_1 and v_2 both being produced by the positive seed generator,

(ii) v_1 and v_2 both being produced by the negative seed generator,

(iii) v_1 being produced by the positive seed generator and v_2 by the negative seed generator,

(iv) v_1 being produced by the negative seed generator and v_2 by the positive seed generator.[3]

The following empirical tests were applied to strings of 1000 standard normal, pseudo-random variates generated by each of the four above methods.

(a) Sign Change Tests with Lags 1 to 25.

These are nonparametric tests of independence based on the count of sign changes between the n^{th} and $(n+i)^{th}$ variates for $n = 1, 2, \ldots, 1000 - i$, with the lag, i, ranging from 1 to 25. The critical values calculated for such two-tail tests are presented in Table A2-1.

(b) Tests for Skewness and Kurtosis

These are the two well known tests for departure from normality based on b_1 and b_2, the sample moment estimates of the coefficient of skewness and the coefficient of kurtosis respectively. Tabulated critical values of $\sqrt{b_1}$ and b_2 for 2 percent and 10 percent two-tail tests published by Pearson and Hartley (1954) were used. Geary

3. Golder and Settle (1976) advocated the use of two unrelated generators, as in (iii) and (iv), in order to counter the possibility raised by Neave (1973) of the innate dependence of uniform deviates causing the resultant normal deviates to have an unsatisfactory distribution.

TABLE A2.1

Critical Values of Count of Sign Changes

for Two-tail Tests on Sample of Size 1000

Lag	1.0%		2.5%		5%		10%	
1	458	541	464	535	468	531	473	526
2	458	540	463	535	468	530	473	525
3	457	540	463	534	467	530	472	525
4	457	539	462	534	467	529	472	524
5	456	539	462	533	466	529	471	524
6	456	538	461	533	466	528	471	523
7	455	538	461	532	465	528	470	523
8	455	537	460	532	465	527	470	522
9	454	537	460	531	464	527	469	522
10	454	536	459	531	464	526	469	521
11	454	535	459	530	463	526	468	521
12	453	535	458	530	463	525	468	520
13	453	534	458	529	462	525	467	520
14	452	534	457	529	462	524	467	519
15	452	533	457	528	461	524	466	519
16	451	533	456	528	461	523	466	518
17	451	532	456	527	460	523	465	518
18	450	532	455	527	460	522	465	517
19	450	531	455	526	459	522	464	517
20	449	531	454	526	459	521	464	516
21	449	530	454	525	458	521	463	516
22	448	530	453	525	458	520	463	515
23	448	529	453	524	457	520	462	515
24	447	529	452	524	457	519	462	514
25	447	528	452	523	456	519	461	514

(1947) has demonstrated that tests based on $\sqrt{b_1}$ or b_2 have optimal large sample properties if the deviation from normality is due solely to skewness or kurtosis respectively.

(c) Chi-squared Goodness-of-Fit Test

An obvious choice of test of the null hypothesis that the variates are independent identically distributed standard normal is the conventional chi-squared goodness-of-fit test. This test was applied by partitioning the real line into 59 equal probability intervals, giving rise to a test statistic whose asymptotic null distribution is chi-squared with 58 degrees of freedom. The choice of 59 intervals was made using Mann and Wald's (1942) formula for determining the optimal number of equal probability cells.

(d) u-test

The empirical results of Shapiro et al. (1968) indicate that David, Hartley and Pearson's (1954) u-test has good sensitivity for detecting symmetric alternative distributions with short tails. This is an important alternative hypothesis in view of the evidence produced by Neave (1973) to support his criticism of the use of the Box-Müller transformation in conjunction with a congruential generator. The test is based on the ratio of sample range to sample standard deviation.

(e) Tests of Minimum and Maximum Values

These two tests for departure from a standard normal random sample are based on the computed probabilities of minimum and maximum variates under the null hypothesis being less than the observed minimum and maximum values respectively. For a two-tail test at the 100α% level, the null hypothesis is rejected if the computed probability is less than $^{\alpha}/_{2}$ or greater than $1-^{\alpha}/_{2}$. Again these tests were chosen because of their obvious potential in detecting short tailed distributions.

(f) Modified Kolmogorov-Smirnov Tests

There are a number of Kolmogorov-Smirnov goodness-of-fit tests. We used the tests of independent identically distributed standard normal variates based on the modified one-sided Kolmogorov-Smirnov one sample statistics, $T(D^{+})$ and $T(D^{-})$ and the modified Kolmogorov-Smirnov one sample statistic $T(D)$ as outlined by Stephens (1970).

(g) D'Agostino's D-test

This is D'Agostino's (1971) large sample extension of Shapiro and Wilk's (1965) w test for departure from normality.

(h) Test of Zero Mean

Under the null hypothesis of a standard normal random sample, the sample mean is distributed $N(0,.001)$. Hence one can test $\sqrt{1000}$ times the sample mean against the standard normal distribution.

(i) Test of Unit Variance

Assuming a standard normal random sample, the sum of squared deviations about the sample mean has a chi-squared distribution with 999 degrees of freedom. This statistic was used to test the null hypothesis that the population variance is unity.

2. RESULTS

Table A2.2 contains the results of the spectral test. The computed values of C_2, C_3, C_4 and C_5 are not test statistics in the strict sense though their values are used to judge the quality of the generator being tested. Based on computational experience, Knuth (1969, p.89) suggested the following two rules of thumb: the generator "passes" the spectral test if C_2, C_3 and C_4 are at least 0.1 and it "passes with flying colours" if C_2, C_3 and C_4 are at least 1. More recently, Hoaglin (1976) expressed the view that "it seems justifiable to insist that any generator intended for serious or public use pass at least the spectral test by having $C_n \geq 1$ for $n = 2, 3, 4, 5, 6$. (One might tolerate a 'near miss' for one or two of these values of n.)"

Both generators pass Hoaglin's requirements for C_2, C_3, C_4 and C_5. Unfortunately Golder's algorithm does not calculate C_6, this being the least important[4] of the C_n statistics for $n = 2, 3, 4, 5, 6$. There is little doubt that the negative seed generator is to be preferred. It compares most favourably with the best of a range of generators tested by Knuth and Hoaglin.

TABLE A2.2

Results of Spectral Test

	C_2	C_3	C_4	C_5
Positive seed generator a = 152587890725	2.9596	0.9164	1.0332	1.4193
Negative seed generator a = 277626315293	2.6906	4.0767	2.4724	6.0913

4. The relative importance of the C_n statistic declines as n increases.

Tables A2.3 - A2.6 contain the results of the empirical testing. Test statistics marked †, *, **, *** are significant at the 10%, 5%, 2.5%, 1% levels, respectively, for two-tail tests.[5] The limited availability of suitable critical values allowed $\sqrt{b_1}$ and b_2 to be tested only at the 10% and 2% levels while the u-test and D'Agostino's D-test were tested at the 10%, 5%, 2% and 1% levels.

The results of the sign change tests hint at possible tenth order autocorrelation in generator (i) although there is conflict as to whether it is positive or negative autocorrelation. Taking this into account, we can find no evidence of unsatisfactory standard normal pseudo-random variates being produced by any of the four methods tested.

3. CHOICE OF GENERATOR

The negative seed generator in conjunction with the Box-Müller transformation was chosen for the generation of standard normal variates in this thesis. This choice was made because of the negative seed generator's spectacular success in passing the spectral test and its ease of application. The empirical testing reported above tends to indicate, that in our particular case, there is little to be gained by using two generators in tandem.

---

5. One-tail tests were applied in the case of the chi-squared goodness-of-fit test.

TABLE A2.3

Results of Empirical Testing on Generator (i)

(v_1 and v_2 produced by positive seed generator)

Seed	2096909	1048931	2097153	524625	524593	1048281	524381	1048647	1048685	2097069
(a) Sign Change Test										
lag 1	511	469†	470†	493	482	508	504	516	525	492
2	506	503	523	504	494	495	503	505	473†	479
3	504	497	518	502	480	490	504	483	501	504
4	514	485	491	503	502	470†	478	479	486	488
5	494	474	511	486	496	513	516	509	493	488
6	521	509	508	491	491	512	497	503	494	499
7	494	505	487	529*	482	501	495	494	503	494
8	494	507	467†	497	473	487	489	498	520	511
9	479	495	497	502	508	493	499	515	501	488
10	460*	524†	512	500	508	518	500	491	526*	461*
11	500	501	467†	513	495	507	483	494	466†	485
12	484	505	514	477	473	482	483	491	485	506
13	490	494	520	485	472	473	505	476	498	516
14	472	526*	496	498	494	500	497	493	491	504
15	500	498	468	469	470	501	505	482	501	472
16	466†	495	509	493	518†	488	491	480	484	482
17	509	508	492	509	501	500	478	510	494	515
18	484	503	480	513	477	487	500	489	475	466
19	484	480	488	475	501	502	470	496	477	476
20	517	457*	502	478	511	503	490	462†	478	496
21	491	485	506	492	501	495	496	480	505	505
22	504	494	502	491	496	466	482	463†	477	499
23	468	513	482	463	496	499	475	411***	501	497
24	468	481	503	492	469	494	485	485	473	489
25	491	479	479	482	475	484	469	479	501	476
(b) Skewness $\sqrt{b_1}$	-.084	-.059	-.056	.079	.153†	-.155†	.042	.014	-.151†	-.032
Kurtosis b_2	3.115	2.861	2.825	3.048	2.925	2.821	2.828	2.853	2.821	3.231
(c) $\chi^2(58)$	55.27	52.91	50.20	56.22	64.48	65.78	50.55	60.47	50.08	49.96
(d) u-Test	6.66	6.37	6.24	6.53	5.93	5.82	6.18	5.97	5.64*	7.77*
(e) Min. Value Probability	.652	.202	.391	.494	.040†	.425	.693	.042†	.228	.9952***
Max. Value Probability	.635	.625	.547	.504	.300	.009**	.072	.341	.045†	.775
(f) Modifed K-S $T(D^+)$	.250	.357	.513	1.163†	.994	.350	1.276*	.895	.588	.415
$T(D^-)$	.857	.552	.397	.328	.469	.504	.126	.229	.345	.546
$T(D)$	.857	.552	.513	1.163	.994	.504	1.276†	.895	.588	.546
(g) D'Agostino D-Test	-.782	1.065	.722	.736	.170	.993	1.059	.710	.965	-.891
(h) Test of Mean Zero	.823	.149	-.179	-.908	-.496	-.206	-1.911*	-.903	-.926	.102
(i) Test of Unit Variance	995.9	964.9	1032.3	957.4	941.4	970.3	995.5	942.5	1025.2	1033.1

TABLE A2.4

Results of Empirical Testing on Generator (ii)

(v_1 and v_2 produced by negative seed generator)

Seed	2096757	524519	1048477	2097149	1048295	2097051	523991	1048863	524367	524453
(a) Sign Change Test										
lag 1	462**	484	500	488	493	468*	519	480	509	512
2	486	509	483	491	513	493	486	482	465*	482
3	501	478	495	488	500	507	509	508	501	475
4	500	462**	497	507	523	486	508	485	463*	494
5	504	485	505	475	528†	465	506	496	507	509
6	477	506	505	510	491	494	497	495	497	500
7	484	500	505	468†	489	503	516	533**	481	509
8	494	524†	493	516	490	501	483	489	470†	486
9	518	487	487	497	486	481	506	485	534**	483
10	499	490	485	489	507	483	484	478	501	441***
11	493	496	512	504	499	463*	517	478	494	493
12	475	488	477	480	520†	491	502	503	520†	490
13	471	508	502	507	505	503	515	497	481	474
14	496	468	487	479	472	502	494	486	488	473
15	481	484	478	477	497	493	480	483	493	488
16	518†	481	516	476	500	508	532**	518†	500	480
17	475	470	488	481	479	500	469	523*	502	475
18	478	474	507	506	496	499	503	490	485	457*
19	492	479	463†	500	498	488	477	516	529**	469
20	503	498	502	495	483	502	491	501	486	498
21	483	484	466	493	477	506	498	502	486	490
22	498	494	489	479	489	475	489	503	479	491
23	512	506	476	472	469	493	511	483	487	497
24	499	485	510	481	480	480	478	479	482	434***
25	502	490	476	481	459†	482	491	495	472	497
(b) Skewness $\sqrt{b_1}$	.118	.027	-.017	-.123	-.025	-.008	-.116	.037	-.046	-.056
Kurtosis b_2	2.923	2.856	3.053	3.088	3.007	2.873	2.919	2.991	3.146	3.054
(c) $\chi^2(58)$	77.93*	59.88	50.67	66.48	52.09	55.86	56.81	58.11	41.47	70.61
(d) u-Test	5.96	6.81	6.86	6.52	6.60	6.59	6.91	6.65	7.34†	6.35
(e) Min. Value Probability	.136	.744	.958†	.888	.658	.805	.894	.472	.921	.813
Max. Value Probability	.270	.801	.021*	.231	.506	.260	.437	.769	.915	.261
(f) Modified K-S $T(D^+)$	.399	1.048	1.168†	.339	.906	.918	.284	.074	.227	.151
$T(D^-)$	.603	.577	.325	.691	.327	.176	.638	.838	.914	1.185†
$T(D)$	.603	1.048	1.168	.691	.906	.918	.638	.838	.914	1.185
(g) D'Agostino D-Test	.111	1.249	-.072	-.601	-.117	.993	.684	-.093	-.595	-.486
(h) Test of Mean Zero	.626	-.879	-1.612	.049	-.858	-.931	.265	1.458	1.388	1.486
(i) Test of Unit Variance	973.9	1039.0	922.8	1038.5	983.1	978.8	978.8	995.7	1046.8	1056.8

TABLE A2.5

Results of Empirical Testing on Generator (iii)

(v_1 produced by positive seed generator, v_2 by negative seed generator)

v_1 Seed / v_2 Seed	1048145 / 524413	524105 / 1048233	524463 / 2096833	2096973 / 524349	1048861 / 1048837	2097075 / 2097055	2097453 / 524541	524151 / 2097371	2097017 / 1048081	1048501 / 2096991
(a) Sign Change Test										
lag 1	475	493	505	515	517	480	511	495	491	499
2	500	528†	499	489	489	494	503	520	538**	505
3	523	489	486	493	520	549***	482	501	493	505
4	500	500	523	497	489	522	519	496	496	491
5	504	476	525†	513	493	528†	502	521	474	473
6	484	520	474	486	504	478	520	489	489	468†
7	485	473	507	495	524†	500	515	487	485	493
8	512	531*	495	536**	507	487	522†	506	486	463*
9	499	518	509	498	523†	498	509	503	497	489
10	503	479	489	506	503	480	490	482	486	492
11	501	486	481	508	504	509	472	497	491	467†
12	505	490	486	505	478	496	476	493	478	497
13	526*	502	480	486	496	505	468	505	494	482
14	496	479	492	475	509	476	475	518	493	483
15	484	469	497	489	485	487	482	489	473	492
16	508	489	470	484	486	499	487	499	503	493
17	485	502	476	488	506	487	491	461†	512	469
18	502	500	492	462	518†	500	506	494	484	491
19	505	491	502	494	505	504	501	502	481	489
20	495	469	473	475	506	498	494	486	512	523*
21	496	463†	492	485	505	492	485	490	510	480
22	463†	511	488	486	517†	489	511	482	490	487
23	453***	494	503	504	475	508	472	504	501	512
24	476	488	479	492	475	471	485	488	508	509
25	464	486	505	492	477	459†	467	473	493	489
(b) Skewness $\sqrt{b_1}$	.034	-.040	-.024	.011	.058	-.013	-.057	-.119	.056	.006
Kurtosis b_2	3.125	2.956	2.990	3.135	3.005	2.815	2.886	3.071	3.147	3.035
(c) $\chi^2(58)$	56.69	57.40	59.88	74.86†	44.30	52.56	59.64	62.24	41.47	47.49
(d) u-Test	7.18	5.71†	6.71	6.28	6.75	5.95	6.07	6.19	6.32	6.10
(e) Min. Value Probability	.121	.167	.929	.342	.118	.108	.478	.185	.251	.205
Max. Value Probability	.987*	.101	.162	.594	.938	.426	.105	.187	.751	.591
(f) Modified K-S $T(D^+)$	.687	1.117	.793	.157	.078	.392	.382	.943	.449	.357
$T(D^-)$	.242	.059	.533	1.256*	.867	.803	.542	.805	.491	.716
$T(D)$	.687	1.117	.793	1.256†	.867	.803	.542	.943	.491	.716
(g) D'Agostino D-Test	-.311	.142	-.020	-1.041	-.035	1.137	.566	-.687	-1.094	-.664
(h) Test of Mean Zero	-.693	-2.438**	.032	1.878†	1.604	.455	.286	-.276	.444	.285
(i) Test of Unit Variance	967.8	1014.0	998.7	1021.0	980.0	1011.5	983.2	898.5*	1036.1	1040.4

TABLE A2.6

Results of Empirical Testing on Generator (iv)

(v_1 produced by negative seed generator, v_2 by positive seed generator)

v_1 Seed	2096845	524619	1048663	2096961	1048279	542599	524635	1048153	2097709	1048047
v_2 Seed	1048517	2097307	524211	524247	1048139	2097157	524389	524337	2096915	1048847
(a) Sign Change Test										
lag 1	493	513	508	504	470†	496	490	500	526†	489
2	487	489	493	508	496	500	508	512	513	500
3	481	492	467*	494	502	489	472†	502	517	515
4	487	498	524†	490	490	499	483	528†	490	501
5	497	492	468†	501	488	494	477	500	520	487
6	483	510	484	504	476	508	510	513	491	496
7	516	496	519	497	490	518	478	482	513	540***
8	505	486	518	477	510	485	509	501	462*	468†
9	500	517	515	467†	480	481	509	488	527*	482
10	515	527*	519	497	494	482	508	459**	472	467†
11	493	486	518	494	478	500	492	464†	501	489
12	488	485	500	491	501	466†	503	507	492	498
13	460*	476	504	492	506	510	490	492	510	484
14	498	498	494	497	473	478	499	490	488	473
15	497	496	501	486	498	478	494	509	509	497
16	482	482	491	480	496	511	504	488	494	470
17	465	490	482	507	465†	489	505	489	473	480
18	530**	503	487	500	493	515	514	485	481	483
19	487	509	493	491	472	509	496	504	512	511
20	493	479	492	481	481	479	495	493	492	479
21	513	506	469	469	490	506	483	486	490	518†
22	484	490	493	495	482	448***	484	455*	466	480
23	488	509	495	488	509	500	500	473	489	485
24	472	503	488	479	500	484	478	475	481	482
25	481	487	477	483	499	481	517†	483	492	494
(b) Skewness $\sqrt{b_1}$	.068	.082	.146†	-.022	.024	.039	-.095	-.035	-.044	-.008
Kurtosis b_2	2.961	2.990	3.045	2.832	2.764	2.857	2.859	2.978	2.966	2.874
(c) $\chi^2(58)$	38.28	44.42	66.48	59.52	55.75	64.48	60.58	61.41	66.25	55.04
(d) u-Test	6.38	6.36	6.69	5.99	6.16	6.48	6.87	6.43	6.05	6.29
(e) Min. Value Probability	.428	.172	.296	.362	.847	.456	.809	.410	.085	.293
Max. Value Probability	.253	.686	.855	.195	.079	.573	.469	.537	.511	.270
(f) Modified K-S $T(D^+)$	.568	.456	.567	.250	1.264*	.471	.336	.405	.915	.313
$T(D^-)$	.697	.543	.717	.659	.725	.841	.994	.711	.078	1.184†
$T(D)$	.697	.543	.717	.659	1.264†	.841	.994	.711	.915	1.184
(g) D'Agostino D-Test	-.134	-.136	-.657	.920	1.414	1.257	1.144	.092	-.218	.542
(h) Test of Mean Zero	.360	.465	.845	1.030	-1.190	.822	1.218	.511	-1.708	1.325
(i) Test of Unit Variance	926.9	977.6	984.2	1011.6	1076.2†	982.3	948.6	975.0	988.9	922.1†